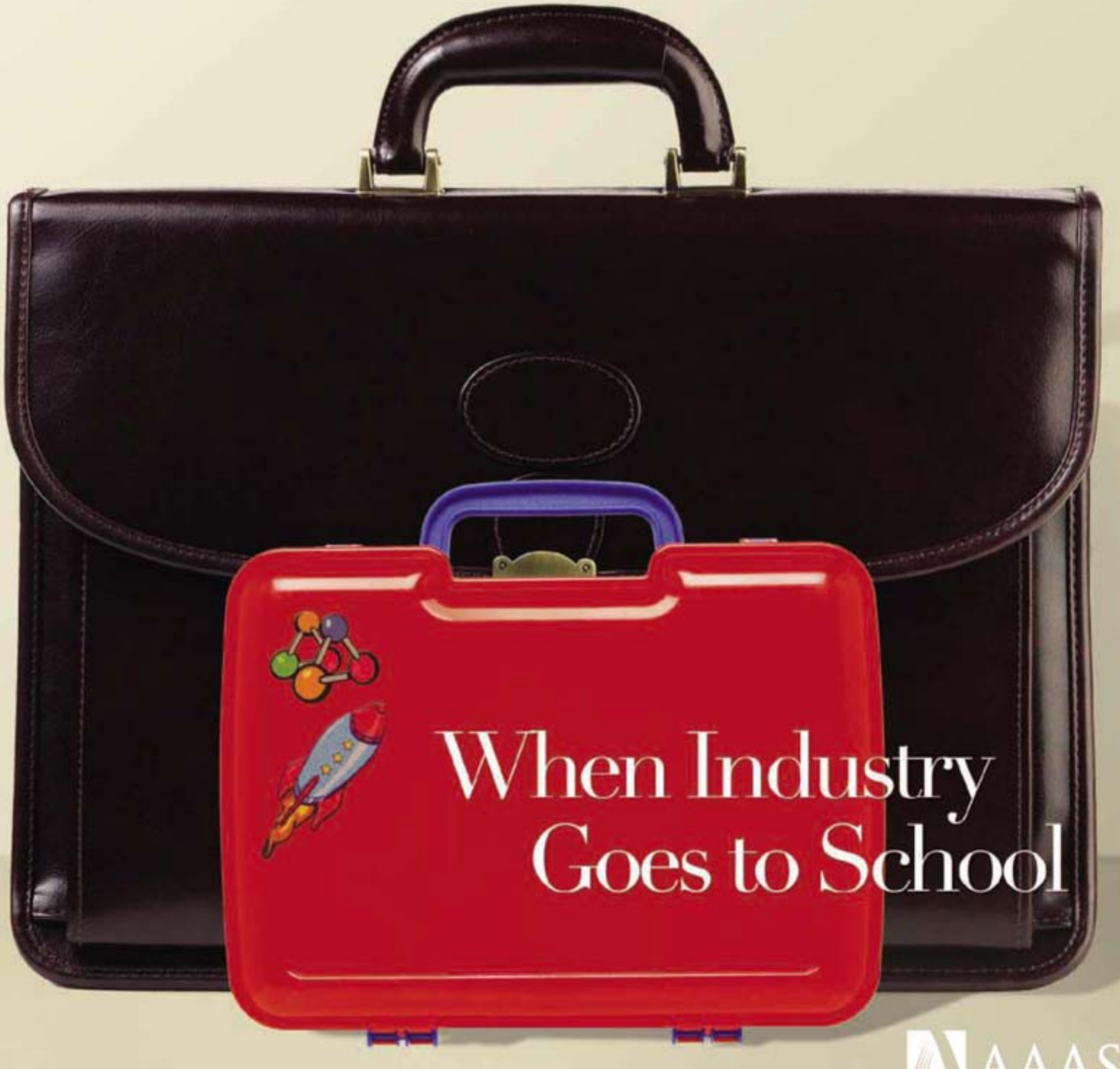
22 February 2008 | S10

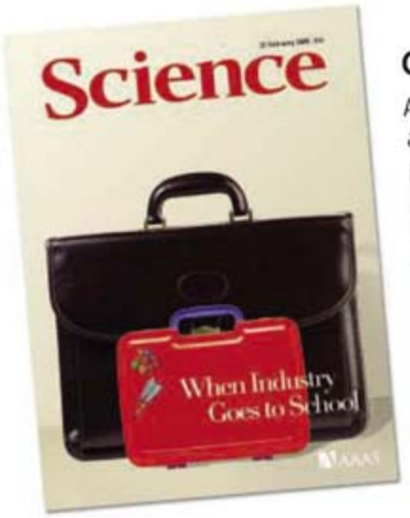
Science



When Industry Goes to School

AAAS

Volume 319, Issue 5866



COVER

A growing number of U.S. corporations are reacting to reports of poor student performance on national and international tests by contributing time and money to efforts aimed at improving math and science education in U.S. schools. To find out what they are doing and how well it's working, see page 1030.

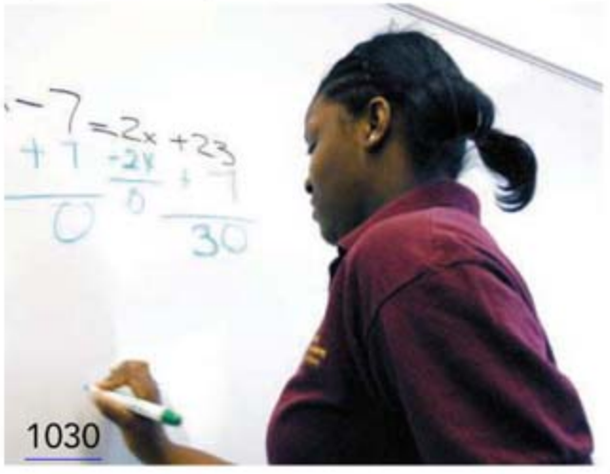
Photo illustration: Kelly Krause/Science (images: Jupiter Images; Getty)

DEPARTMENTS

EDITORIAL

NEWS OF THE WEEK

NEWS FOCUS



LETTERS

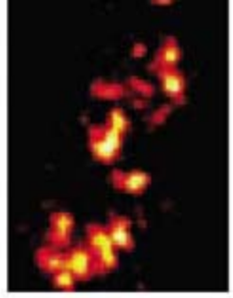
BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

CONTENTS continued >>

Science

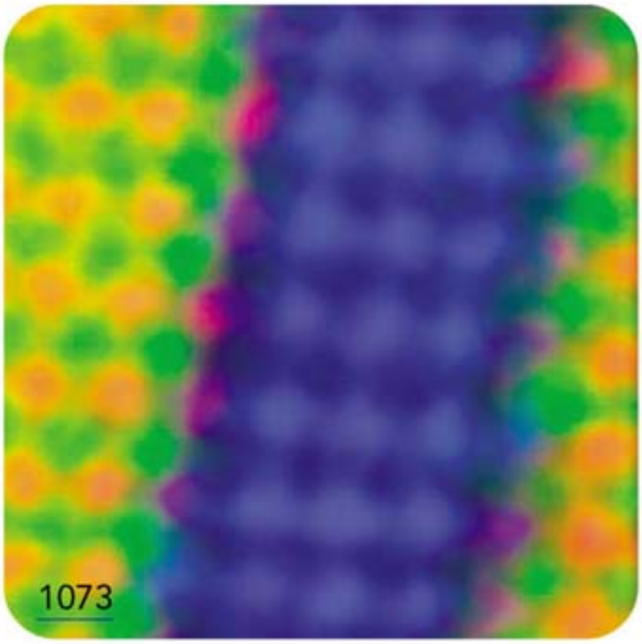


SCIENCE EXPRESS

www.sciencexpress.org

TECHNICAL COMMENT ABSTRACTS

REVIEW



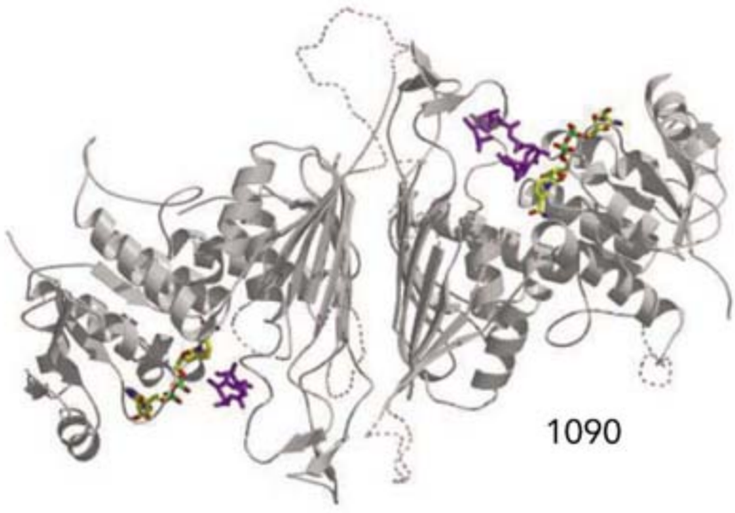
1073

BREVIA

REPORTS

CONTENTS continued >>

REPORTS *CONTINUED...*



1090

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership and subscription (51 issues): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $770; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. SCIENCE is printed on 30 percent post-consumer recycled paper. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CONTENTS continued >>

The purple sea urchin is headed for trouble.

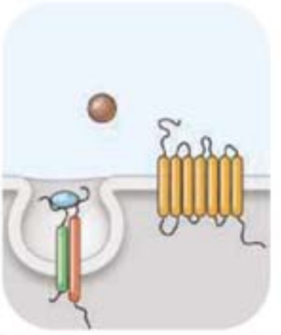
SCIENCE**NOW**

www.sciencenow.org DAILY NEWS COVERAGE

Warm Sea Urchins on Acid
Rising temperatures and acidity in the ocean cause deformities and death.

A Grand Diversion in Louisiana
Model predicts how much wetlands would benefit from shunting Mississippi River.

Anticipation Fires the Imagination
New research helps explain why finding new love—and other experiences—don't always meet expectations.



Kremen regulates Wnt signaling.

SCIENCE **SIGNALING**

www.stke.org THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: Context-Dependent Activation or Inhibition of Wnt–β-Catenin Signaling by Kremen
C. S. Cselenyi and E. Lee
The effect of Kremen 2 on Wnt signaling depends on the presence or absence of the Wnt antagonist Dickkopf1.

PERSPECTIVE: A Cytoskeletal Platform for Local Translation in Axons
F. P. G. Van Horck and C. E. Holt
Mutual interactions between the cytoskeleton and local translation may mediate growth cone steering response.

PODCAST
N. R. Gough and J. F. Foley
Listen to a discussion about the signals that generate fat cells.

Benefiting from life's setbacks.

SCIENCE **CAREERS**

www.sciencecareers.org CAREER RESOURCES FOR SCIENTISTS

Opportunities: Master of Disaster
P. Fiske
Personal and professional setbacks can open the door to new opportunities.

Mastering Your Ph.D.: Preparing for Your Post-Ph.D. Career
B. Noordam and P. Gosling
Find out how to find a job that suits you, inside or outside academia.

Alone in the Field
E. Pain
Spain's science funding agency rejected Eduardo Moreno's idea, but it still won him a grant from the European Research Council.

***Science* Careers Blog**
Science *Careers Staff*
Get the latest career news, including reports from the AAAS annual meeting in Boston.

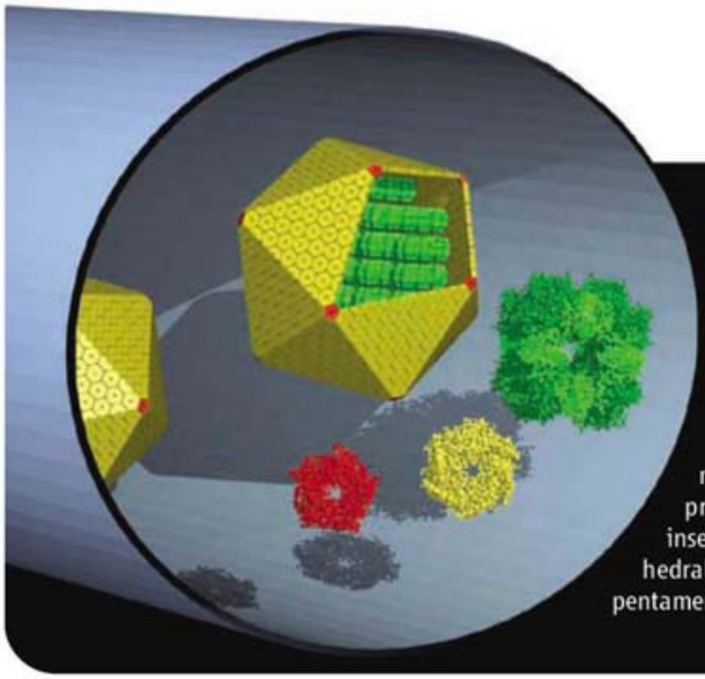
SCIENCE**PODCAST**

Download the 22 February *Science* Podcast to hear about insights into human genetic diversity, corporate support of math and science education, the force required to move an atom, and more.
www.sciencemag.org/about/podcast.dtl

CREDITS: (SCIENCE NOW) DAVID MONNIAUX; (SCIENCE CAREERS) BEN OSTROWSKY/CREATIVE COMMONS

Separate individual or institutional subscriptions to these products may be required for full-text access.

EDITED BY STELLA HURTLEY AND PHIL SZUROMI



<< Carboxysome Building Blocks

The carboxysome is a bacterial microcompartment, roughly 100 nanometers in diameter that sequesters enzymes involved in carbon fixation. The proteinaceous outer shell is icosohedral, a geometry that typically involves a combination of hexameric and pentameric building blocks such as those found in certain viral capsids. Four abundant shell proteins from two known types of carboxysomes are known to form hexamers. Now **Tanaka *et al.*** (p. 1083) report the structure of two carboxysome proteins that form pentamers with the appropriate size and shape to be inserted into hexagonally packed layers so that they form vertices in an icosohedral shell. Model-building gave two plausible ways to pack the hexamers and pentamers to build a complete carboxysome shell.

Ductility in Ceramic Composites

Nature has found routes to take weak or brittle starting materials and make stronger or tougher composites, such as incorporating biopolymers and calcium carbonate to form nacre. Although scientists have been able to use better starting materials, they have not been able to achieve the same level of improvement or sophistication. **Bonderer *et al.*** (p. 1069; see the Perspective by **Ortiz and Boyce**) fabricate a composite from submicrometer-thick alumina embedded in a polymer matrix (chitosan) that has a high tensile strength. By controlling the thickness of the ceramic films, they can determine the mode of fracture and ensure that the composites remain ductile.

Photon Turnstile

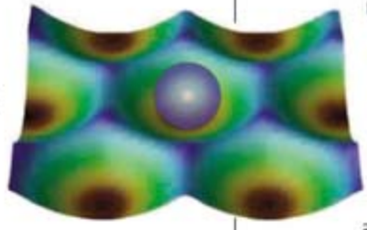
High-quality optical microresonators can influence or delay the propagation of the light in coupled optical fibers. **Dayan *et al.*** (p. 1062) now show that coupling of an atom to the evanescent field of the resonator can have a more dramatic effect on the propagating light in the fiber. When the atom is coupled, the input stream of photons is regularized into an output of single photons. Such an optical turnstile should prove useful as a light source for quantum information processing.

Whither Weather Out West

Recently, more winter rain and less snow has fallen in the western United States than previously, and more snow has melted earlier in the year. These changes have increased river flows in the spring and decreased them in the summer. **Barnett *et al.*** (p. 1080, published online 31 January) conducted a multivariate climate change detection and attribution study of the western United States from 1950 to 1999 and found that up to 60% of the trends of river flow, winter air temperature, and snow pack reflect anthropogenic climate change despite high variability throughout the records. These results suggest that a crisis in water supply for the western United States, which relies heavily on a late spring snow melt for agriculture, may be looming.

A Little Nudge

Scanning probe microscopes have been used to manipulate individual atoms and create structures at the atomic scale, but despite their widespread use, it has been difficult to determine the forces needed to move each atom. **Ternes *et al.*** (p. 1066; see the Perspective by **Custance and Morita**) report on simultaneous measurements of the vertical and lateral forces exerted on an adsorbate, either Co atoms or CO molecules, during lateral manipulation using an atomic force microscope (AFM). They simultaneously detected the tunneling current between the oscillating AFM tip and the sample surface, which recorded variations of the tunneling gap conductance with the force measurements.



Social Quakes

Earthquakes constantly color Earth's plate boundaries, but how and to what extent are any two quakes connected, let alone a series of quakes? Clear aftershocks cover only a small number of quakes and primarily series that are successively weaker. Overall, the question is partly statistical, and partly mechanistic in determining how one quake affects stress on another fault. **Marsan *et al.*** (p. 1076) develop a model of triggering based on two simple assumptions without the need for fitting a large number of parameters for each fault. Their model, applied to southern California quakes, implies that small earthquakes have a disproportionate effect on triggering and that cascading events are common.

Tau and Microtubule Motors

The microtubule-associated protein tau is implicated in neurodegenerative diseases, including Alzheimer's disease, but neither its normal nor pathological role are well understood. Microtubules in cells are densely coated with microtubule-associated proteins, including tau, and **Dixit *et al.*** (p. 1086) wanted to learn how motor proteins navigate on coated microtubules. The authors incorporated several isoforms and truncations of tau in a cell free assay to build a complex system mimicking the cell environment. Encounters between individual motor proteins and tau on microtubules were observed directly, which revealed dramatic differences between the effects of tau on the motility of two microtubule-based motors, dynein versus kinesin. It appears likely that tau can spatially regulate the balance of microtubule-dependent axonal transport in neurons.

Continued on page 1007

CREDITS (TOP TO BOTTOM): T. YEATES; TERNES *ET AL.*

Donald Kennedy is the Editor-in-Chief of *Science*.

Confidential Review—or Not?

AT *SCIENCE*, WE EDITORS LOVE OUR REVIEWERS AND KNOW THAT OUR EDITORIAL colleagues elsewhere do too. After all, the process of scientific publication depends on the volunteer services of thousands of experts all over the world who willingly provide, without compensation, confidential and candid evaluations of the work of others. Because all of us in scientific publishing depend on reviewers, we'd better try to keep them at it, happy, and secure. But the following case, involving a lawsuit, a drug company, and the company's assault on the confidential files of a journal, is a bad news story.

The drug company Pfizer is being sued in various jurisdictions on product liability grounds. Plaintiffs are claiming that its products Celebrex and Bextra cause cardiovascular and other injuries. Pfizer asserts that in some cases plaintiffs are making use of published papers from the *New England Journal of Medicine* (*NEJM*). So it wants to dig though the confidential reviews of those papers in search of something to strengthen its defense. The company served *NEJM* with a series of subpoenas to which the journal replied, claiming several privileges in support of its refusal to comply.

Now Pfizer's lawyers have filed a motion to compel *NEJM* to produce the files, which will be heard by a U.S. District Court in Massachusetts. (Full disclosure: I have filed an affidavit with the court supporting *NEJM*.) The motion is interesting in terms of its revelations about what Pfizer knows about the process of scientific publication and what it regards as the "public interest." For example, the motion states: "The public has no interest in protecting the editorial process of a scientific journal . . ." Say what? Doesn't the public want access to credible biomedical science? If not, what was the open-access movement all about? Do medical advocacy groups really have no use for knowledge that might help their members?

Does confidentiality count for anything to the scientists who serve the journal? Well, if confidentiality is compromised, Pfizer's attorneys state with breezy assurance, that won't be a problem for authors: "It is unreasonable to conclude," they say in their motion to compel, "that scientists and academics will stop submitting manuscripts to *NEJM* if it complies with this subpoena." Perhaps. But what about reviewers, who are explicitly promised confidentiality? And what about other journals? If this motion succeeds, what journal will not then become an attractive target for a similar assault?

Viewed in the larger context, this is really a conflict between competing interests. One is the public's interest in a fair system of evaluating and publishing scientific work—one that offers high confidence in, though not an absolute guarantee of, the quality of the product. Pfizer dismisses this with a wave of the hand, a strangely inconsistent position given the enthusiasm with which it and other drug companies seek to have their own research validated by the very system of scientific publication that Pfizer's motion decries and would undermine. On the other side, there is a private interest in gaining information that might protect a corporate defendant against a plaintiff's attack. Without questioning the legitimacy of the latter, it is surely fair to ask whether fulfilling that need should trump the public interest.

An approach often taken in such cases would examine the prospective weight of what defendant Pfizer hopes to find; in other words, is it worth it? What Pfizer's motion says on that score is: "Scientific journals such as *NEJM* may have received manuscripts that contain exonerating data for Celebrex and Bextra which would be relevant for Pfizer's causation defense." That's a pretty frank admission that this is a fishing expedition in which Pfizer hopes it "may" find something to help its defense by exposing a reviewer's comment. Is that an adequate basis for justifying prospective damage to the public interest? We don't think so, and we suspect our prospective reviewers won't think so, either. But if efforts of this kind were to succeed, the sad day might come when *Science* would have to add a firm caveat emptor to its instructions for peer reviewers.

– Donald Kennedy

10.1126/science.1156250

EDITED BY GILBERT CHIN AND JAKE YESTON

EVOLUTION

Under Strict Control

In many species of animals, males and females are strikingly (and visibly) different. Three evolutionary genetic pathways can produce such sexual dimorphism. From the outset, the expression of the alleles for dimorphic traits could be limited to one sex. Or the alleles could be expressed initially in both sexes but subsequently promoted only in one sex or repressed only in the other. Darwin believed that the bright colors and ornaments of males were male-limited from their first appearance, whereas Fisher held that they had been lost from females through the accumulation of suppressors. Coyne *et al.* weigh these alternatives by examining female hybrids between bird species [such as Costa's (left) and black-chinned (right) hummingbirds] whose males show distinct ornamental traits (for example, plumage patterns and the presence or absence of elongated feathers). Unless the suppressors are completely dominant, these females should show some expression of the male traits. In none of the 13 hybrid cases (from four orders and eight families) that met their search criteria did the females express a male-limited trait. This suggests that male ornaments were sex-limited at their first appearance (perhaps by hormonal or physiological differences between the sexes) or are now controlled by completely dominant modifier alleles or regulatory elements. — SJS

Evolution **62**, 214 (2008).

PLANETARY SCIENCE

Telling Time on Mars

Our primary information on the surface age and geologic history of Mars comes from crater densities on different parts of the planet and from dating of Martian meteorites (though their source region on Mars is not known with certainty). Crater ages imply that most of the crust is over 2 billion years old. In contrast, dating of one distinct group of Martian meteorites, by a variety of isotopic systems, has generally indicated young ages, from a few hundred thousand to about 1 billion years. These numbers imply that Mars had enough recent magmatic activity to be sampled by a significant fraction of the available Martian meteorites. A problem, however, is that all of these rocks show evidence of alteration and impact events (required for but not limited to their excavation from Mars), as well as hints of the very early geochemical history of Mars, so interpreting their ages is not generally straightforward. Bouvier *et al.* present additional analyses of two of the purportedly young meteorites using several isotopic systems and evaluate the ages of all of the Martian meteorites. In general, the Sm-Nd and Lu-Hf ages derived from minerals were young, but the Pb-Pb whole-rock ages were much older (~4 billion years). In contrast to some other recent analyses, the authors interpret this difference, and ultimately the other young ages, as indicating later alteration that has re-initialized mineral-based isochrons. — BH

Earth Planet. Sci. Lett. **266**, 105 (2008).

APPLIED PHYSICS

Plasmons on a Wire

The miniaturization of optical elements such as waveguides is limited by the wavelength of light, typically from a few hundred nanometers up to the micrometer scale. By comparison, electrical circuits can be patterned down to a few tens of nanometers. A promising route for integrating these two scales is direct conversion of an optical signal into an electrical one via surface plasmons created at the surface of metal dielectric interfaces. Schmidt *et al.* use a photonic crystal fiber as the optical carrier and show that they can fill in the air holes that run the length of the fiber with gold or silver, introduced in molten form, to create an array of metal nanowires embedded within the silica matrix. Characterization of the hybrid structure shows that plasmon resonances are excited in the nanowires, thereby demonstrating the direct integration of optics and nanoelectronics. — ISO

Phys. Rev. B **77**, 33417 (2008).

PLANT SCIENCE

Resisting Resistance

Bacillus thuringiensis (*Bt*) toxins are bacterial proteins that kill some insects and are widely used in crops for pest control. Plants expressing the *Bt* toxins Cry1Ac and Cry1Ab make up the majority of insect-resistant transgenic corn and cotton, respectively. A major concern has been that the pests will develop resistance; refuges, where non-transgenic plants grow next to transgenic plants, have been hypothesized to delay the evolution of resistance because resistant individuals will mate with susceptible individuals from the refuge and hence not confer resistance (if inheritance is recessive) to their offspring.

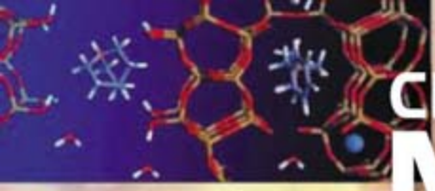
The cotton pest *Helicoverpa zea*.

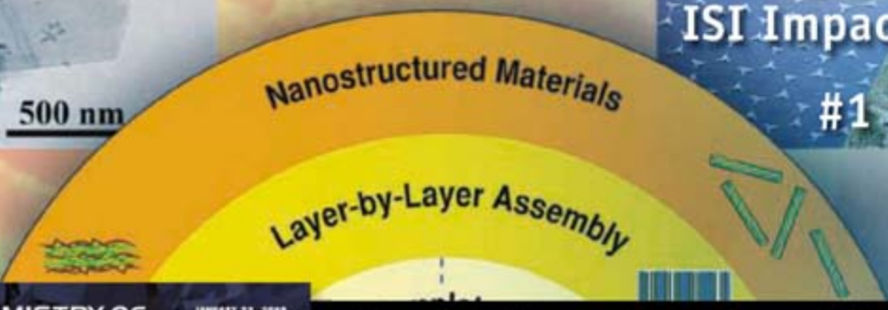
Tabashnik *et al.* have analyzed field studies from around the world to determine the effectiveness of the refuge strategy. They found that resistance has evolved in populations of the corn earworm at some U.S. locations, but not elsewhere nor in five other pests. Furthermore they conclude that the refuge strategy has been successful in slowing the evolution of resistance and that expressing multi-

Continued on page 1013

CREDITS (TOP LEFT AND RIGHT): RICK & NORA BOWERS/ALAMY; (BOTTOM): SCOTT BAUER/USDA

CHEMISTRY OF MATERIALS

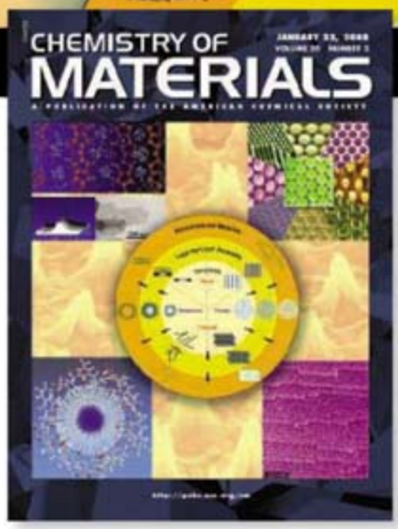




ISI Impact Factor: 5.104

#1 most-cited in materials science

http://pubs.acs.org/CM



Volume 20, Issue 3
February 12, 2008

Editor-in-Chief: Leonard V. Interrante
Rensselaer Polytechnic Institute

20TH 1989-2008 ANNIVERSARY

To view a sample issue of *Chemistry of Materials* and lists of most-accessed and most-cited articles, go to http://pubs.acs.org/CM

Special Issue: Templated Materials

Templating is one of the most versatile methods for the creation of structured materials, resulting in structure sizes typically between nanometers and micrometers, but even extending into the macroscopic world. Structures can be one-dimensional, such as in the case of nanotubes, two-dimensional, such as patterned surfaces, or three-dimensional, as found in zeolites, ordered mesoporous materials, or colloidal crystals. This variety of templated structures finds its correspondence in the types of templates used, which range from molecules over supramolecular arrays to polymer latices and others. Applications of templated materials are manifold: zeolites, as one of the most important class of technical catalysts, are an example of a well-established technology, while colloidal crystals are possible key elements in emerging applications in photonic devices.

Editors:

Ferdi Schüth
Max-Planck-Institut für Kohlenforschung

Mietek Jaroniec
Kent State University

This Special Issue on "Templated Materials" provides a survey of the current topics and major lines of development in this rapidly growing research area. It covers a broad spectrum of templated materials with a variety of possible applications. Due to the high current interest in this field and the exciting developments of the last decade, such a Special Issue is very timely indeed.

Authors of Review Articles to this Special Issue:

Markus Antonietti, *Max-Planck-Institute of Colloids and Interfaces*
Carl Batt, *Cornell University*
Frank Caruso, *University of Melbourne*
Jun Chen, *Nankai University*
Claus Christensen, *Danish Technical University*
Francisco del Monte, *Instituto de Cienca de Materiales de Madrid*
Kazunari Domen, *University of Tokyo*
Allan Guymon, *University of Iowa*
Marc Hillmyer, *University of Minnesota*
Shinji Inagaki, *Toyota Research Labs*
Plinio Innocenci, *University di Sassari*
Susumu Kitagawa, *Kyoto University*
Yue Li, *AIST Tsukuba*
Klaus Mosbach, *Lund University*
Clement Sanchez, *Université Pierre et Marie Curie Paris*
Andreas Stein, *University of Minnesota*
Toshimi Shimizu, *AIST Tsukuba*
Hyunjung Shin, *Kookmin University*
Galen Stucky, *University of California at Santa Barbara*
Takashi Tatsumi, *Tokyo Institute of Technology*
Michael Tiemann, *Universität Giessen*
Dongyuan Zhao, *Fudan University*

Continued from page 1011

ple *Bt* toxins in single plants may further slow the appearance of resistance. — LMZ

Nat. Biotechnol. **26**, 199 (2008).

CHEMISTRY

Air to the Rescue

Chemists generally take pains to ensure the purity of their reagents, lest a highly reactive contaminant doom the anticipated outcome. Once in a while, though, an adventitious impurity proves more help than hindrance. The reported synthesis of quinine by Woodward and Doering in 1944 was a landmark achievement in the history of organic chemistry, but in the ensuing years, controversy swelled over the report's reliance on a final reduction step published 26 years earlier by Rabe and Kindler. Was the chemistry in this last step really viable? Had Woodward taken his synthesis all the way to fruition, or actually stopped at the penultimate step? In a meticulous effort to reproduce the Rabe study, Smith and Williams now find that modern concerns probably arose on account of excessively pure materials. They show that the reduction proceeds most efficiently when the aluminum powder called for in the published procedure has been partially oxidized by exposure to air, a reasonably likely circumstance in 1918. — JSY

Al powder.

Angew. Chem. Int. Ed. **47**, 1736 (2008).

IMMUNOLOGY

Exceptional Acceptance

In 1960, Medawar and Burnet were awarded the Nobel Prize for Physiology or Medicine for showing that immunological tolerance to foreign grafts can be acquired. What this and many studies in animal models have shown since is that it is possible to "re-educate" the immune system of the recipient to the foreign antigens carried by the donor cells and grafts. The induction of full immune tolerance to transplanted organs has also been pursued in the clinic because this would dispense with the need for extensive immunosuppressive regimes currently used to prevent organ rejection.

Scandling *et al.* and Kawai *et al.* report cases of human kidney transplantation in which hematopoietic cells from the donor were co-transplanted with the organ, followed by distinct protocols of conditioning involving depletion of host cells. In the first study, this resulted in ongoing mixed chimerism; meaning that there were comparable proportions of donor and recipient hematopoietic cells circulating in the body. In the second study, the extent of chimerism was considerably less and was also transient. Nevertheless, in both clinical settings it was possible to discontinue immunosuppressive drugs without significant adverse responses to the grafted organs; furthermore, cellular evidence of immunological tolerance was detected. In a third study, Alexander *et al.* describe a female liver transplant patient in which a spontaneous switch occurred in the blood group and—because the donor was male—the sex of the circulating leukocytes. Impressively, the removal of immunosuppressive drugs, which was undertaken to resolve hemolytic anemia in the patient, resulted in complete chimerism (meaning that all hematopoietic cells were of donor origin) and long-term acceptance of the grafted liver. These three studies reveal that immunological acceptance of transplanted organs is a realistic goal worthy of further exploration. — SJS

N. Engl. J. Med. **358**, 362; 353; 369 (2008).

PHYSIOLOGY

We're Never Alone

Each of us carries a community of microorganisms in our intestinal tract that is necessary for optimal health but varies in composition from person to person. To learn which gut microbes most influence human physiology, Li *et al.* surveyed these symbionts in seven members of a four-generation Chinese family. By sequencing rRNA genes, they assembled a phylogenetic picture of the resident microbes—mainly *Bacteroidetes* and *Firmicutes* spp. The Chinese microbiomes showed individual and male/female differences in their composition, although all differed markedly from those seen in Americans. A parallel analysis of urine samples from the same family allowed the authors to correlate variations in resident microbes with variations in excreted metabolites, which served as a proxy for the individuals' metabolic state. They found several microbial species that appeared to significantly affect their hosts: for instance, *Faecalibacterium prausnitzii* was associated with the presence of dimethylamine, suggested to be an indicator of metabolic syndrome and diabetes. This functional metagenomics approach can be extended by combining full genomic sequencing of the microbes with measurements of metabolic or clinical features of interest. Ultimately, the hope is to identify which gut inhabitants are associated with disease states and which with human well-being. — KK

Proc. Natl. Acad. Sci. U.S.A. **105**, 2117 (2008).

CREDIT: 2008 THEODORE GRAY

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store http://www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services http://www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org (for general editorial queries)
science_letters@aaas.org (for queries about letters)
science_reviews@aaas.org (for returning manuscript reviews)
science_bookrevs@aaas.org (for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science and innovation throughout the world for the benefit of all people. The goals of the association are to: foster communication among scientists, engineers and the public; enhance international cooperation in science and its applications; promote the responsible conduct and use of science and technology; foster education in science and technology for everyone; enhance the science and technology workforce and infrastructure; increase public understanding and appreciation of science and technology; and strengthen support for the science and technology enterprise.

Information for Authors

See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Donald Kennedy**
EXECUTIVE EDITOR **Monica M. Bradford**
DEPUTY EDITORS **R. Brooks Hanson, Barbara R. Jasny, Katrina L. Kelner**
NEWS EDITOR **Colin Norman**

Editorial supervisory senior editor Phillip D. Szuromi; **senior editor/perspectives** Lisa D. Chong; **senior editors** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Valda Vinson, David Voss; **associate editors** Jake S. Yeston, Laura M. Zahn; **online editor** Stewart Wills; **associate online editors** Robert Frederick, Tara S. Marathe; **book review editor** Sherman J. Suter; **associate letters editor** Jennifer Sills; **editorial manager** Cara Tate; **senior copy editors** Jeffrey E. Cook, Cynthia Howe, Harry Jach, Barbara P. Ordway, Trista Wagoner; **copy editors** Lauren Kmec, Peter Mooreside; **editorial coordinators** Carolyn Kyle, Beverly Shields; **publications assistants** Ramatoulaye Diop, Chris Filiatreau, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Scott Miller, Jerry Richardson, Brian White, Anita Wynn; **editorial assistants** Carlos L. Durham, Emily Guise, Patricia M. Moore, Jennifer A. Seibert; **executive assistant** Sylvia S. Kihara; **administrative support** Maryrose Madrid

News senior correspondent Jean Marx; **deputy news editors** Robert Coontz, Eliot Marshall, Jeffrey Mervis, Leslie Roberts; **contributing editors** Elizabeth Culotta, Polly Shulman; **news writers** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin, David Grimm, Constance Holden, Jocelyn Kaiser, Richard A. Kerr, Eli Kintisch, Andrew Lawler (New England), Greg Miller, Elizabeth Pennisi, Robert F. Service (Pacific NW), Erik Stokstad; **intern** Elsa Youngsteadt; **contributing correspondents** Jon Cohen (San Diego, CA), Daniel Ferber, Ann Gibbons, Robert Irion, Mitch Leslie, Charles C. Mann, Virginia Morell, Evelyn Strauss, Gary Taubes; **copy editors** Rachel Curran, Linda B. Felaco, Melvin Gatling; **administrative support** Scherraine Mack, Fannie Groom; **bureaus** New England: 207-549-7755, San Diego, CA: 760-942-3252, FAX 760-942-4979, Pacific Northwest: 503-963-1940

Production director James Landry; **senior manager** Wendy K. Shank; **assistant manager** Rebecca Doshi; **senior specialists** Jay Covert, Chris Redwood; **specialist** Steve Forrester; **Preflight director** David M. Tompkins; **manager** Marcus Spiegler; **specialist** Jessie Mudjitaba

Art director Kelly Buckheit Krause; **associate art director** Aaron Morales; **illustrators** Chris Bickel, Katharine Sutliff; **senior art associates** Holly Bishop, Laura Creveling, Preston Huey, Nayomi Kevitiyagala; **associate** Jessica Newfield; **photo editor** Leslie Blizard

***Science* International**

Europe (science@science-int.co.uk) **editorial: international managing editor** Andrew M. Sugden; **senior editor/perspectives** Julia Fahrenkamp-Uppenbrink; **senior editors** Caroline Ash, Stella M. Hurtley, Ian S. Osborne, Stephen J. Simpson, Peter Stern; **editorial support** Deborah Dennison, Rachel Roberts, Alice Whaley; **administrative support** John Cannell, Janet Clements, Jill White; **news: europe news editor** John Travis; **deputy news editor** Daniel Clery; **contributing correspondents** Michael Balter (Paris), John Bohannon (Vienna), Martin Enserink (Amsterdam and Paris), Gretchen Vogel (Berlin); **intern** Elizabeth Quill

Asia Japan Office: Asca Corporation, Eiko Ishioka, Fusako Tamura, 1-8-13, Hirano-cho, Chuo-ku, Osaka-shi, Osaka, 541-0046 Japan; +81 (0) 6 6202 6272, FAX +81 (0) 6 6202 6271; asca@os.gulf.or.jp; **asia news editor** Richard Stone (Beijing: rstone@aaas.org); **contributing correspondents** Dennis Normile (Japan: +81 (0) 3 3391 0630, FAX 81 (0) 3 5936 3531; dnormile@gol.com); Hao Xin (China: + 86 (0) 10 6307 4439 or 6307 3676, FAX +86 (0) 10 6307 4358; cindyhao@gmail.com); Pallava Bagla (South Asia: +91 (0) 11 2271 2896; pbagla@vsnl.com)

Africa Robert Koenig (contributing correspondent, rob.koenig@gmail.com)

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner**

Fulfillment Systems and Operations (membership@aaas.org); **director** Waylon Butler; **customer service supervisor** Pat Butler; **specialists** Laurie Baker, Latoya Casteel, LaVonda Crawford, Vicki Linton; **data entry supervisor** Cynthia Johnson; **specialists** Tarrika Hill, Erin Layne, Sheila Thomas

Business Operations and Administration director Deborah Rivera-Wienhold; **assistant director, business operations** Randy Yi; **senior financial analysts** Michael LoBue, Jessica Tierney; **financial analyst** Nicole Nicholson; **rights and permissions: administrator** Emilie David; **associate** Elizabeth Sandler; **marketing director** John Meyers; **marketing managers** Allison Pritchard, Darryl Walter; **marketing associates** Aimee Aponte, Alison Chandler, Mary Ellen Crowley, Marcia Leach, Julianne Wielga, Wendy Wise; **international marketing manager** Wendy Sturley; **marketing executive** Jennifer Reeves; **marketing/member services executive** Linda Rusk; **site license sales director** Tom Ryan; **sales manager** Russ Edra; **sales and customer service** Mehan Dossani, Iquo Edim, Kiki Forsythe, Catherine Holland, Phillip Smith, Philip Tsolakidis; **electronic media: manager** Lizabeth Harman; **project manager** Trista Snyder; **assistant manager** Lisa Stanford; **senior production specialist** Walter Jones; **production specialists** Nichele Johnston, Kimberly Oster

Advertising director worldwide ad sales Bill Moran

Product (science_advertising@aaas.org); **consumer & sponsorship sales manager** Tina Morra: 202-326-6542; **midwest** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **west coast/w. canada** Teola Young: 650-964-2266; **east coast/e. canada** Christopher Breslin: 443-512-0330, FAX 443-512-0331; **uk/europe/asia** Michelle Field: +44 (0) 1223-326-524, FAX +44 (0) 1223-325-532; **japan** Mashy Yoshikawa: +81 (0) 33235 5961, FAX +81 (0) 33235 5852; **senior traffic associate** Deiandra Simms

Commercial Editor Sean Sanders: 202-326-6430

Classified (advertise@sciencecareers.org); **us: recruitment sales manager** Ian King: 202-326-6528, FAX 202-289-6742; **inside sales manager: midwest/canada** Daryl Anderson: 202-326-6543; **northeast** Alexis Fleming: 202-326-6578; **southeast** Tina Burks: 202-326-6577; **west** Nicholas Hintibidze: 202-326-6533; **sales coordinators** Erika Foard, Rohan Edmonson, Shirley Young; **international: sales manager** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **sales** Marium Hudda, Alex Palmer, Alessandra Sorgente; **sales assistant** Louise Moore; **japan** Mashy Yoshikawa +81 (0)3 3235 5961, FAX +81 (0)3 3235 5852; **advertising production operations manager** Deborah Tompkins; **senior production specialists** Robert Buck, Amy Hardcastle; **senior traffic associate** Christine Hall; **publications assistant** Mary Lagnaoui

AAAS Board of Directors **retiring president, chair** John P. Holdren; **president** David Baltimore; **president-elect** James J. McCarthy; **treasurer** David E. Shaw; **chief executive officer** Alan I. Leshner; **board** John E. Dowling, Lynn W. Enquist, Susan M. Fitzpatrick, Alice Gast, Linda P. B. Katehi, Cherry A. Murray, Thomas D. Pollard, Kathryn D. Sullivan

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Naming Rights

NASA's Gamma-ray Large Area Space Telescope (GLAST) may be scientists' best hope for studying gamma-ray bursts throughout the universe. But the name doesn't exactly roll off the tongue. So NASA is asking the public to come up with something better.

The $690 million observatory will aid researchers studying gamma-ray bursts—brief, intense flashes of light as bright as a million trillion suns. The bursts are thought to be linked to many celestial phenomena, including supernovae and black holes.

GLAST, scheduled for a May launch, is a follow-up to NASA's Compton Gamma Ray Observatory, which was operational from 1991 to 2000. NASA officials say the new observatory possesses a stronger sensitivity to gamma rays and a wider view of the universe than its predecessor.

The deadline to submit names is 31 March, and the winning name will be announced once the observatory is in orbit and all systems are tested. To suggest a name, visit the NASA Web site at glast.sonoma.edu/glastname/.

Shiny, Plastic People

You may already be a part of Ken and Barbie's plastic family. A new U.S. government study suggests that the bodies of most Americans are laced with bisphenol A (BPA), the primary component of one of the most widely used plastics.

BPA is an ingredient in polycarbonate plastics, used in everything from baby bottles to composite dental fillings to toy dolls. Frequent use and heating can cause the plastic to break down, leaching BPA into whatever the plastic touches, including food. Although recent studies have linked BPA to reproductive problems in lab animals, little is known about humans and BPA.

To measure BPA exposure, chemist Antonia Calafat and colleagues at the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, examined urine samples from 2517 participants in the ongoing National Health and Nutrition Examination Survey. Significant levels of BPA were present in 92% of the samples. Those with the highest levels were children and adolescents, women, and low-income participants.

The body metabolizes BPA in only a few hours, suggesting that exposure to BPA is continuous, says epidemiologist Russ Hauser of Harvard School of Public Health in Boston. Although the jury's still out on what high exposures mean for human health, Hauser says industry should consider alternatives that aren't biologically active.

MOLECULAR NITPICKING

Christopher Columbus may not have brought the epidemic typhus that ravaged the New World after his arrival. In fact, the parasites may predate him by tens of thousands of years, according to a new study.

Typhus, a potentially deadly flulike disease, is spread by body lice. David Reed, an evolutionary biologist at the Florida Museum of Natural History in Gainesville, and Didier Raoult of the University of the Mediterranean in Marseille, France, sequenced DNA from ancient lice preserved in the scalps and braids of 1000-year-old Peruvian mummies. To their surprise, the researchers discovered that the lice belong to a subtype found all over the world.

Although this subtype typically infests the head, it also includes body lice, which thrive in clothing and spread diseases such as epidemic typhus. Thus, "it's reasonable to believe that body lice were in the New World" well before Columbus, says Dale Clayton, an evolutionary biologist at the University of Utah, Salt Lake City.

Deciphering ancient DNA isn't new; researchers have sequenced DNA from extinct mammoths and Neandertals. But DNA in lice tends to degrade quickly, and the oldest lice genetic material sequenced to date comes from Napoleon's soldiers—bodies buried 200 years ago. With this new study, "it's as though a parasitologist or epidemiologist traveled 1000 years back in time to sample a parasite from a human patient," says Mark Hafner of Louisiana State University in Baton Rouge.

Virtual Dissection

Dissecting a formalin-soaked frog isn't part of most middle-school science curricula anymore. But a new software program promises to keep the sense of discovery—and the frog—alive. The virtual-reality program, called V-Frog, allows students to dissect internal organs, explore tissues using an endoscope, and even observe a beating heart—all with the click of a mouse.

Computer-based multimedia dissections aren't new, as schools seek to avoid the use of chemicals and sharp objects as well as the controversy over the use of animals. But most programs are not interactive and offer only a few predetermined options.

The new program was developed by computer scientist Kevin Chugh at Tactus Technologies near Buffalo, New York. Based on photographs and artistic renderings, V-Frog runs on a standard Windows-based personal computer. Using interactive and 3D technology, students can manipulate tissues so that each dissection is different. "It's intuitive for them," he says. "It's the teachers who need the manual."

Jayne Mackta, president of the New Jersey Association for Biomedical Research in Union, says that virtual frogs enhance the learning experience but don't duplicate traditional dissections. "Do they completely eliminate the need for a real-model organism? I'd like to think no."

CREDITS (TOP TO BOTTOM): NASA AND GENERAL DYNAMICS; TACTUS TECHNOLOGIES; DR. SONIA GUILLEN, CENTRO MALLQUI, PERU

Three Q's

As a clinician-researcher studying autism at the University of Washington, Seattle, **Geraldine Dawson**, 56, spent many years helping parents cope with afflicted children. Last month, Dawson became the first science adviser for Autism Speaks, whose $33-million-a-year research budget makes it the largest funder of autism research in the world.

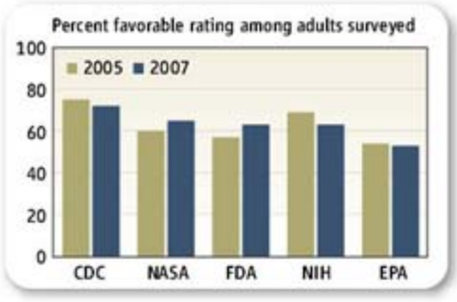
Q: What's hot in autism research?
There's so much in genetics. And we're starting to translate some of the findings into animal models. We're understanding brain pathology better. There's strong interest in trying to attract pharma companies, and we're trying to facilitate that by starting a clinical-trials network that will allow us to be nimble.

Q: What can you offer parents in the near-term?
Many of these kids suffer from anxiety, seizures, gastrointestinal problems. The clinical-trial network will try to develop drugs for core symptoms. And if we discover more genes and biomarkers, these are going to allow us to pick up kids at risk very early. We still hope that by beginning [behavioral] interventions before age 1, because of brain plasticity, the outcomes will be more positive.

Q: Many studies have ruled out vaccines as a cause of autism, but some parents demand more research. Do you fund any?
There are questions that haven't been addressed, such as the impact of having so many vaccines in such a short time for a genetically vulnerable child. We don't see this as the primary focus [for our organization] but one area among many that are important. My role is to listen to all the voices and ultimately design studies to come up with answers.

INSIDE GOVERNMENT

SAGGING STOCK. The U.S. National Institutes of Health (NIH) has lost favor among the U.S. public in the past 2 years, according to a new public opinion survey that asked 2000 adults how they felt about 22 federal agencies. Among the five science agencies included in the survey, NIH showed the biggest decline: 63% of those surveyed held a favorable opinion of the agency, down 6 percentage points since 2005.



Despite public concerns about the safety of certain medications on the market, the Food and Drug Administration's rating rose 6 percentage points, to 63%. NIH spokesperson John Burklow says that NIH will continue efforts to educate the public about the value of biomedical research.

IN BRIEF

Cancer biologist **Inder Verma** of the Salk Institute for Biological Studies in San Diego, California, has won a $50,000 prize from the New York City–based Vilcek Foundation to honor foreign-born Americans. Verma, who was born in India and studied at both the University of Lucknow and the Weizmann Institute of Science in Rehovot, Israel, before joining Salk in 1974, is being recognized for his contributions to understanding the genetic alterations that cause cancer and developing techniques for gene therapy.

A geneticist at the University of Pittsburgh has been sentenced to 1 year of unsupervised probation and a $500 fine for shipping microbes to a Buffalo, New York, artist (*Science*, 9 July 2004, p. 159). **Robert Ferrell**, 64, pleaded guilty to a misdemeanor in October after sending Steven Kurtz the microbes—*Serratia marcescens and Bacillus atrophaeus*—for use in an art exhibit. Many scientists were outraged that the U.S. government prosecuted Ferrell, especially because the organisms are normally harmless. "We're very pleased" that the case is over, says Efrem Grail, Ferrell's lawyer. The case against Kurtz is ongoing.

Got a tip for this page? E-mail people@aaas.org

Movers >>

BRIEF TERM. David Schwartz has resigned as director of the National Institute of Environmental Health Sciences (NIEHS), ending a stormy 3-year tenure. The 54-year-old pulmonologist drew criticism for attempting to privatize the institute's journal and shift funds from disease prevention to clinical studies. But a bigger blow came from an inquiry by Congress last year suggesting that he ran afoul of ethics rules when he consulted for law firms and built up a large personal lab. Schwartz claimed that any errors were the result of misunderstandings, but he temporarily stepped down as director in August and took up an advisory role at the U.S. National Institutes of Health (NIH) while the agency reviewed NIEHS management (*Science*, 24 August 2007, p. 1021).

His 8 February resignation brings the chapter to a close from his perspective, although NIH's review, which has yet to be completed, is expected to touch on his controversial actions. In an e-mail to his staff, Schwartz explained that NIEHS "would be more successful with new leadership." In addition, Schwartz wrote that he had "inadvertently disenfranchised segments of our community," for which "I sincerely apologize."

Schwartz will join the National Jewish Medical and Research Center in Denver, Colorado, where he will head a new genetics research center and direct the pulmonary and critical-care division.

CREDITS (TOP TO BOTTOM): AP; GFK ROPER REPORTS; NIH

Counting down India's tigers
1027

Fresh from the AAAS meeting
1028

No treatment. Elderly Chinese patients suffering from dementia play chess to keep mentally active.

GENETICS

Once Shunned, Test for Alzheimer's Risk Headed to Market

A Pennsylvania company is preparing to market a genetic test that will tell healthy people whether they are at increased risk for developing Alzheimer's disease. The move comes more than 15 years after the critical gene, *APOE*, was linked to Alzheimer's, and it is getting a mixed response from researchers. Some of them point out that the test could upset people without giving any therapeutic benefit. On the other hand, as the company says, the information has its uses, and research has shown that receiving a bad result is not as devastating as once feared.

The test will be offered by Smart Genetics in Philadelphia, likely starting next month. For $399, healthy people will give a saliva sample and learn whether they have a risk of Alzheimer's that's 3 to 15 times higher than normal. The analysis is based on variations in the *APOE* gene, which is widely agreed to play a role in Alzheimer's risk and heart disease.

The test differs from many other gene tests for common adult diseases. For one, the science behind it is solid. And the psychological ramifications are under study: REVEAL, a large clinical trial funded by the U.S. National Institutes of Health in Bethesda, Maryland, has been examining how healthy people react to learning their *APOE* genotype and how best to communicate this potentially explosive information. Smart Genetics executives say that they're building on the findings. They plan to screen out those who seem emotionally unstable and provide a genetic counseling session by telephone before giving out *APOE* results.

Still, researchers express reservations about making the gene test widely available. They worry about the mental health consequences of telling people they may get a disease that's neither preventable nor treatable and is invariably fatal. "I think the benefits [of knowing your genotype] are trivial" and don't justify the emotional risks, says law professor Henry Greely of Stanford University in Palo Alto, California, who in 1997 co-chaired a working group on Alzheimer's genetic testing. The group concluded that genetic testing for Alzheimer's "is not appropriate for most people."

> *"It isn't helpful if there's nothing you can do about it."*
>
> —ALLEN ROSES, DUKE UNIVERSITY

Greely says his views have shifted only slightly in favor of testing. He thinks knowing the results might help the roughly 2% of the population with the worst *APOE* combination: two copies of the deleterious E4 allele, which together confer a roughly 15 times increased risk of the disease. For them, Greely says, the risk is so great that the information may be useful in planning health care needs or retirement.

But a much larger portion of the population, about 25%, carries one copy of *APOE4*; their risk of Alzheimer's is roughly three times higher than normal. Greely doesn't think these people need to know their *APOE* status, and Allen Roses of Duke University in Durham, North Carolina, who found the *APOE*-Alzheimer's link, agrees: "It isn't helpful if there's nothing you can do about it" medically.

But Richard Watson, chief technical officer of Smart Genetics, argues that knowing one is at higher risk can trigger practical responses. Watson says these might include regular memory screenings or making certain financial decisions such as buying long-term care insurance.

Until now, the only company offering an *APOE* gene test has been Athena Diagnostics in Worcester, Massachusetts, and the only patients who qualify for testing have been those with dementia. Athena licensed rights to the gene patent from Duke University, where Roses made the *APOE* discovery in 1992. Smart Genetics has in turn licensed rights to the test from Athena.

Some felt that an *APOE* test could be useful to people without symptoms. In 1999, neurologist Robert Green of Boston University and his colleagues launched a study to examine how healthy people with a family history of Alzheimer's responded to learning their *APOE* genotype. The news wasn't harmful for the 500 or so who've received the information, Green's group found. But at this point, "we don't fully know if disclosing *APOE* is safe psychologically," says Green, who has offered unpaid advice to Smart Genetics.

Of course, wider use of the *APOE* gene test might turn up adverse effects. But the REVEAL study has provided some reassurances to those who worried that people getting *APOE* genotypes would "freak out, they'd go commit suicide, something really bad would happen," says Robert Cook-

CREDIT: CHINA PHOTOS/GETTY IMAGES

The calculus of corporate giving
1030

The evolution of Stephen Godfrey
1034

Deegan of the Duke Institute for Genome Sciences and Policy in Durham. The data from REVEAL, he says, suggest that "people who don't get tested at all and people who get bad news stay in more or less the same trajectory." Nonetheless, a person's *APOE* status may be considered particularly sensitive information, Cook-Deegan says. It's the only genetic information that James Watson, the DNA discoverer who recently had his entire genome sequenced, kept secret.

Smart Genetics says it has hired enough genetic counselors to handle a large volume of tests. (The company declines to say how many.) "We saw there was a big growth" in genetic testing and believed "there was something there for adding value to what people wanted," says Julian Awad, the company's CEO. Smart Genetics, he says, is also pondering whether to tell clients about how *APOE4* can raise their risk of heart disease.

Back in the 1990s, says Awad, "the test itself was before its time." Now, he believes, people are ready. **–JENNIFER COUZIN**

U.S. BUDGET

House Panel Berates Science Adviser on 2009 'Shortfall'

Congressional Democrats and Republicans don't agree on much. But last week, members of the House Committee on Science and Technology complained in one voice that President George W. Bush has fallen short on his promise to bolster U.S. innovation in his 2009 budget request (*Science*, 8 February, p. 714).

"You've proposed more of the same in science education and in energy research," Representative Bart Gordon (D–TN), chair of the committee, scolded John Marburger, the president's science adviser and sole witness at the Valentine's Day hearing. "And what you've heard today from both sides of the aisle is that more of the same is not getting the job done."

At issue was how the 2009 budget request stacks up against the America COMPETES Act, a set of goals for science funding that was passed overwhelmingly by Congress last summer and signed by the president (*Science*, 10 August 2007, p. 736). It authorizes spending of $43 billion over 3 years to increase the country's pool of scientific talent and boost research spending. These are also the top two recommendations of the U.S. National Academies' influential 2005 report, *Rising Above the Gathering Storm*.

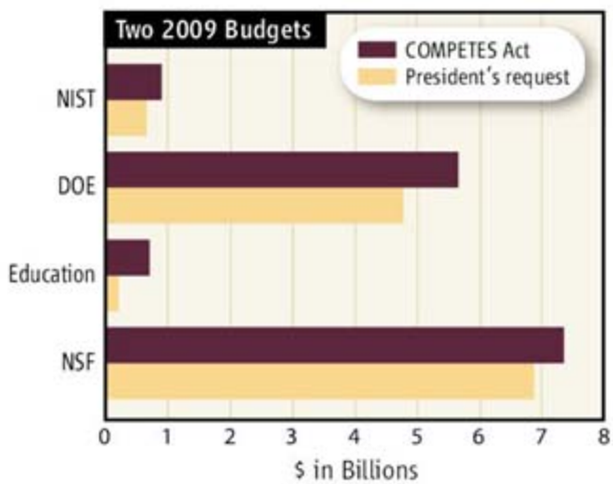
The Bush Administration has embraced the report's second priority with a 10-year plan to double the budgets of the National Science Foundation (NSF), the Department of Energy's (DOE's) Office of Science, and the National Institute of Standards and Technology (NIST) in a package it calls the American Competitiveness Initiative (ACI). But science committee members say ACI falls short on aid for precollege math and science education. They also criticize the Administration for failing to request funding for a DOE entity, the Advanced Research Projects Agency–Energy (ARPA-E), created by the new law.

Adding to their unhappiness was the last-minute loss of large increases slated for the three ACI agencies in 2008—a step reluctantly taken by Democratic leaders after Bush insisted on cuts in overall domestic spending (*Science*, 4 January, p. 18). Science lobbyists are hoping Congress will make a $500 million apology as part of an upcoming supplemental spending bill to finance the Iraq war. But the chances appear slim.

In the meantime, committee members were not buying Marburger's arguments that the Administration's new budget is a good-faith effort to approach the 2009 spending levels in the COMPETES Act. Although the budget proposes spending 84% of the authorized level for DOE science, 93% for NSF, and 71% for NIST, it falls short of the overall target by $2.1 billion (see graphic).

Some of the harshest criticism came from Republicans. "I'm wondering if you're fighting hard enough for your share of the pie," asked Representative Phil Gingrey (R–GA), before hastening to add, "although as a fiscal conservative, I want to keep that pie as small as possible." Representative Vern Ehlers (R–MI) demanded to know "why didn't the Administration propose funding the ACI agencies at the '09 authorization levels?" And Representative Roscoe Bartlett (R–MD) dismissed the Administration's proposed DOE research portfolio, claiming that "nothing you have mentioned in energy research is cutting-edge."

Marburger loyally defended Bush's budget. He told Ehlers that the proposed 20% boost for DOE science and NSF research "represents a huge increase," especially when overall domestic spending is essentially held flat. "Increased funding for critical basic research in the physical sciences is my highest budget priority," he said. "And I think our priorities are correct."



Promises, promises. The authorized spending levels in the America COMPETES Act far exceed what the president has requested for 2009.

SOURCE: OSTP

But Gordon saw things differently. Giving the Department of Education the lead role over NSF for improving math and science education "ensures that we'll continue our mediocre performance on international tests," he said, and refusing to fund ARPA-E makes it less likely that the type of advances needed to develop new energy sources will occur. Gordon says he'll fight hard for changes in the 2009 budget. But his best shot may come after a new president takes office in January. "This is not a partisan issue. And I know that Dr. Marburger is only the messenger."

–JEFFREY MERVIS

SCHOLARLY PUBLISHING

Harvard Faculty Votes to Make Open Access Its Default Mode

Harvard University has jumped into the contentious debate on open-access publishing with a plan to make research papers freely accessible online. The 12 February vote by its 730-member arts and sciences faculty marks the first time that a major U.S. university has directly challenged the authority of academic journals to control access to research results. "This is a large and very important step for scholars throughout the country," said Harvard computer scientist Stuart Shieber, who championed the plan. But the decision does not apply to the rest of the university, including its medical, public health, and business schools, and publishers say they doubt it will significantly affect their business.

This is the university's "first step in the creation of an open-access environment."

—STEVEN HYMAN, HARVARD UNIVERSITY

The resolution authorizes Harvard to place a faculty member's work in a repository that will be available to all at no cost. The researcher would retain the copyright, as in the past, but the university would have a license to release it. Papers would be posted on the Internet upon receipt or following a requested delay. Any author may choose not to participate but would need a waiver from the dean in order to opt out of the system. "What's new is that this is not imposed by the university or by funders," says John Wilbanks, vice president of Science Commons, a Cambridge, Massachusetts, organization supporting open access. "This is the faculty stepping up and saying it has a stake."

"Publishers may not be quite as excited to take articles from Harvard."

—PATRICIA SCHROEDER, ASSOCIATION OF AMERICAN PUBLISHERS

The vote comes 1 month after the U.S. National Institutes of Health required all grantees to place peer-reviewed papers in the agency's free archive, PubMed Central, within a year of publication. The Harvard decision, by contrast, "is not mandatory," notes Patricia Schroeder, president of the Association of American Publishers in Washington, D.C. "I don't think anyone is quaking in their boots." Alan Leshner, chief executive officer of the American Association for the Advancement of Science (which publishes *Science*), says that the new policy won't affect publishing criteria, although it could pose "a bureaucratic problem for faculty members."

Shieber sees the new archive as a way to combat rising subscription costs and give authors greater control over their own published work. Provost Steven Hyman, who supports the change, sees it as the university's "first step in the creation of an open-access environment for current research." Other private organizations—including the Howard Hughes Medical Institute in Chevy Chase, Maryland, and the Wellcome Trust in London—have taken similar steps (*Science*, 7 July 2006, p. 29).

Publishers maintain that the cost of peer review, printing, and distribution require them to charge subscription fees. And they note that small nonprofit organizations, as well as large corporations, profit from the scholarly publishing system and depend on journal subscriptions to stay in business.

Schroeder says it is too early to measure the impact of the new policy and warns that "publishers may not be quite as excited to take articles from Harvard." Shieber agrees that the experiment is just beginning but adds that "journals will always try to get the best papers."

–ANDREW LAWLER

CREDIT (TOP TO BOTTOM): HARVARD; ASSOCIATION OF AMERICAN PUBLISHERS

Candidates Invited to Debate

The first-ever U.S. presidential science debate will take place 18 April in Philadelphia—if any of the four major candidates still in the race agree to come. The organizers, which include Nobelists, university presidents, the U.S. National Academies, and AAAS (the publisher of *Science*), hope the event (www.sciencedebate2008.org) will highlight the role that science plays in a range of national policy issues, from health care and energy independence to education and climate change. The campaigns have yet to respond to the invitations, which went out last week, but Barack Obama is giving it "serious consideration" says an adviser.

–ELI KINTISCH

Kernel of Truth

Corn growers are about to have their decade-long dream of a deciphered maize genome come true. Just in time for the 50th annual Maize Genetics meeting in Washington, D.C., next week, Richard Wilson of Washington University in St. Louis, Missouri, and his colleagues have pieced together a rough draft of corn's 2.3 billion bases, almost a year ahead of schedule. Corn jump-started the National Science Foundation's (NSF's) Plant Genome Initiative, abetted by Senator Christopher "Kit" Bond (R–MO), who formerly chaired the panel that oversees NSF's budget. Already, researchers are experimenting with using the genome to improve crop yields and harness corn's potential as a biofuel, says conference organizer Thomas Brutnell of Cornell University.

–ELIZABETH PENNISI

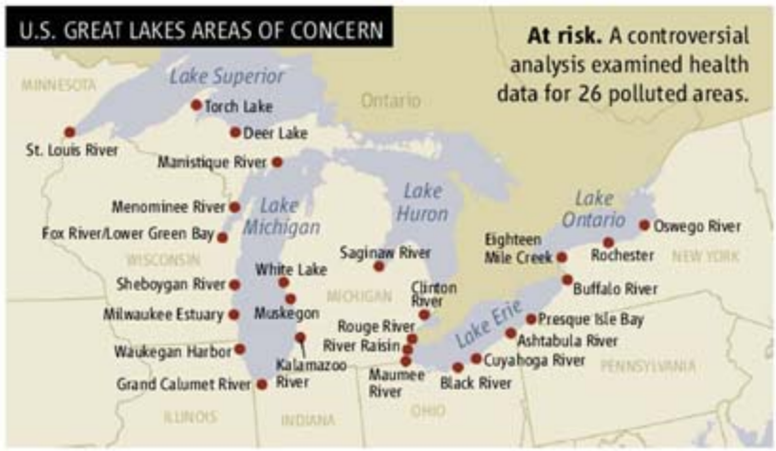
Censorship Rules Slow to Come

In the wake of several cases of apparent scientific censorship by U.S. government officials, Congress last summer ordered President George W. Bush to adopt government-wide rules ensuring that federal scientists are free to disseminate their research results without "suppression or distortion." Last week, presidential science adviser John Marburger said at a hearing of the House Science and Technology Committee that a draft document will soon be circulated among the relevant agencies. There will be no opportunity for public comment, Marburger acknowledged under questioning. The White House is running late—the policy was supposed to be in place by 9 February. "Better late than never" was the reaction of committee chairman, Representative Bart Gordon (D–TN). "This has never been a priority for them." –JEFFREY MERVIS

ENVIRONMENTAL HEALTH

Lawmakers Claim Great Lakes Report Was 'Suppressed'

Lawmakers are again asserting that the Bush Administration is meddling in science. House Science Committee Democrats charge that federal officials have suppressed a report on potential health threats from pollution in the Great Lakes. They also say officials at the U.S. Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, may have punished a career federal scientist who oversaw it. CDC says the report had genuine scientific flaws.



At risk. A controversial analysis examined health data for 26 polluted areas.

The controversy involves Christopher De Rosa, who from 1992 until last fall headed the toxicology division of CDC's Agency for Toxic Substances and Disease Registry (ATSDR). His troubles began after the International Joint Commission, a U.S.–Canadian organization that oversees issues involving the countries' boundary waters, asked ATSDR in 2001 to examine the human health threat from Great Lakes pollution. The region has 26 well-known "areas of concern" with high levels of PCBs, mercury, dioxins, and other contaminants.

Under De Rosa's direction, ATSDR scientists gathered publicly available census and health data for the polluted areas. After peer review by three academics and scores of government scientists, the final draft was completed last February. A copy leaked to the nonprofit Center for Public Integrity in Washington, D.C., concludes that more than 9 million people reside in the areas of concern, and many live in counties that have elevated rates of cancer, low birth weight, and other health problems. Although the report "makes no causal inferences or associations," it says risks in these counties "merit further attention."

When De Rosa planned to release the report in July, ATSDR Director Howard Frumkin and his deputy objected, canceling a press conference. Instead, they sent copies to past peer reviewers, noting among other "concerns" that the report relied on health statistics from before 1998 and limited exposure data. ATSDR asked reviewers whether the report should even be released.

Two outside peer reviewers told the agency it should be. "These criticisms [by Frumkin] are valid," says Peter Orris, an occupational health researcher at the University of Illinois, Chicago. But however flawed, he says, the information "would allow people to say, 'Maybe there's a relationship here' and set a research agenda." Like Orris, environmental health researcher David Carpenter of the University at Albany in New York state suspects that "someone did not want this information to get attention." CDC spokesperson Glen Nowak says CDC expects to submit a revised draft to the Institute of Medicine or another group for review within a few weeks.

Meanwhile, the House Science Committee has rushed to De Rosa's defense. In a 6 February letter to CDC Director Julie Gerberding, chair Bart Gordon (D–TN) and other Democrats say CDC "may have retaliated against" De Rosa. Last fall, after De Rosa sent Frumkin a letter protesting that he had "opposed the release of information to the public on several important health issues," including the cancer risks of formaldehyde in trailers issued to Hurricane Katrina refugees, De Rosa received a poor performance review. He was reassigned to a special assistant position. Nowak declined to comment, saying it's a personnel matter. The latest example of the suppression of science by the Bush Administration? De Rosa says that "is for others to decide."

–JOCELYN KAISER

CREDITS (TOP TO BOTTOM): ADAPTED FROM EPA; KAREN TWEEDY-HOLMES/POPULATION COUNCIL INC.

AIDS RESEARCH

Microbicide Fails to Protect Against HIV

HIV prevention research suffered another setback this week with the failure of the largest trial yet of a microbicide used by women to prevent sexual transmission of the virus. The microbicide, a gel called Carraguard that women insert before having intercourse, contains carrageenan, a seaweed derivative that thwarts HIV in test-tube studies. More than 6000 women at three sites in South Africa participated in the $40 million, placebo-controlled trial, which was funded by the Bill and Melinda Gates Foundation and the U.S. Agency for International Development. "Carraguard was shown to be safe but not effective against HIV," said principal investigator Khatija Ahmed, a microbiologist at the University of Limpopo in South Africa.

In a 14 February teleconference, Ahmed reported a statistically insignificant difference in infection rates: 151 women in the placebo group versus 134 who received Carraguard. Women said they used the gel only 44.1% of the time, and just 10% said they always used it before sex. "That overall number is low, and it could have had an impact on our final study results," said Barbara Friedland, a project manager for the study at the nonprofit Population Council in New York City, which developed the product. But the low usage reflects a real-world scenario, she noted.

Didn't gel. Carraguard failed to live up to its name, providing women no protection against HIV.

Several researchers who study anti-HIV microbicides were disappointed but not surprised by the results. Last year, a large study of a similar product, cellulose sulfate, ended

SPECIES CONSERVATION

Tigers in Decline, Indian Survey Finds

NEW DELHI, INDIA—The use of new sampling techniques has cut by half the estimated number of wild tigers in India. A new report this week from the Indian government puts the number at 1411, compared with 3642 in 2002. Experts say the decline reflects more than just a change in methodology: Poaching, human encroachment, and habitat loss take a heavy toll.

"The tiger is in a state of crisis; there is no denying that fact," says Rajesh Gopal, member secretary for the National Tiger Conservation Authority of the Ministry of Environment and Forests in New Delhi. Gopal dismissed as "complete trash" the previous estimates, based primarily on tallying tiger "pugmarks," or footprints.

CREDIT: P. BAGLA/*SCIENCE*

Predator's predator. A police raid in Ghaziabad highlighted the toll poaching takes on India's tigers.

Officials at the Wildlife Institute of India in Dehradun, which conducted the new study, adopted the new sampling methods 4 years ago after heavy criticism of the reliability of pugmarks (*Science*, 30 May 2003, p. 1355). The present survey deployed 88,000 observers across India, matching their findings with satellite imagery to estimate tiger densities in each of India's 28 states. The "unprecedented effort" cost $2 million, says lead author Yadvendradev V. Jhala, and will be repeated every 4 years.

The new approach used camera-trap surveys, line-transect sampling, and occupancy modeling, says ecologist Melvin Sunquist of the Department of Wildlife Ecology and Conservation at the University of Florida, Gainesville. However, it omitted a vast marshy area known as the Sundarbans forest for which researchers had not developed a sampling method. The margin of error was 17%.

Not everyone endorses the findings. The chief minister of the eastern Indian state of Orissa disputes the survey's estimate of 45 wild tigers in his state, which claimed 173. But Gopal stands by the report, saying, "There is just no prey base to support such large numbers as [Orissa] has claimed."

John Seidensticker, a large-mammal specialist and head of the Conservation Ecology Center at the Smithsonian Institution's National Zoological Park in Washington, D.C., says, "India should be applauded for moving to a science-based framework monitoring their tiger populations." And Sunquist thinks these animals can still be conserved in the wild provided the Indian government makes "a major renewed commitment to the protection of tiger habitat, tiger prey, and the tiger itself."

–PALLAVA BAGLA

when an interim analysis showed more infections in treated women than in the placebo group. (Both cellulose sulfate and Carraguard do not attack HIV directly but have negative charges that interfere with its binding to immune cells.) An earlier trial of nonoxynol-9, which also has nonspecific activity against HIV, actually increased a woman's chance of becoming infected. For AIDS researcher Kenneth Mayer of Brown University, these results mean "the field should now focus on microbicides that are specific against HIV."

A dozen microbicides in human studies contain various anti-HIV drugs. "We'll need the force of several different protocols looking at different products to really answer the questions," says Sharon Hillier of the University of Pittsburgh School of Medicine, who heads the Microbicide Trials Network funded by the U.S. National Institutes of Health. She adds that the need for women to insert a microbicide right before sex contributes to low use. This fall, Hillier plans to compare whether daily use of an anti-HIV drug works better in a microbicide or as an oral pill. "From contraceptives we've learned that women would rather take a pill every day, even if they're only having sex four times a month," she says.

Mayer says Carraguard's failure shouldn't be "the death knell" of microbicides. "We just have to be humble and move forward," he says. "The stakes are too high to do anything else."

–JON COHEN

Ocean Carbon Scheme Sunk

Citing financial woes, a U.S. firm that wanted to sell carbon credits for ocean iron fertilization has said it will "indefinitely postpone" its efforts.

Last month, Planktos, a company based in Foster City, California, began an Atlantic cruise with its Weatherbird II vessel to study using iron to catalyze the growth of plankton. Planktos hoped such algal blooms would take in carbon dioxide and fall to the deep ocean, sequestering the carbon for centuries (*Science*, 30 November 2007, p. 1368). But some prominent scientists questioned whether Planktos could properly monitor the experiment, and Greenpeace and a number of ocean groups concerned about ecological side effects petitioned international bodies to shut down the group. Planktos kept secret details of the cruise, citing violent threats to its ship. Last week, company officials said "a highly effective disinformation campaign … has caused the company to encounter serious difficulty in raising the capital needed." "On top of the very legitimate scientific and technological concerns, the added secrecy didn't help their case at all," says marine biologist David Santillo of Greenpeace.

–ELI KINTISCH

No Plans for British Astronauts

LONDON, U.K.—Britain will not be getting its own space agency or training astronauts any time soon, the government said last week. Its new civil space strategy for 2008–2012 explains that more study is needed before committing to the huge expense of sending people into space.

Various U.K. agencies already spend roughly $430 million per year on projects coordinated by the British National Space Centre (BNSC). Last year, the Royal Society and others urged the government to transform BNSC into a space agency with a budget to fund space missions (*Science*, 27 April 2007, p. 531), arguing that such a step would give it a higher profile and more clout in international collaborations. The new strategy has no boost in funds for new scientific missions and promises only that BNSC will review its own structure. Science minister Ian Pearson says that now is not a good time politically to pitch for a space agency. A BNSC review of exploration, including human space flight, is due out within a year.

The strategy "is strong on its vision, … yet there are few signs of the action required to deliver," says Martin Rees, Astronomer Royal and president of the Royal Society.

–DANIEL CLERY

Not like us. Even smart animals are "whoppingly different" from humans, a researcher asserts.

How Human Intelligence Evolved—Is It Science or 'Paleofantasy'?

BOSTON, MASSACHUSETTS—Richard Lewontin knows how to grab an audience's attention. Lewontin, an evolutionary biologist at Harvard University, led off a session titled "The Mind of a Toolmaker" by announcing that scientists know next to nothing about how humans got so smart. "We are missing the fossil record of human cognition," Lewontin said at the meeting. "So we make up stories."

FINDINGS
The Science Magazine News Blog

For more coverage of the AAAS meeting, see **blogs.sciencemag.org/newsblog**

Not so, responded other human evolution experts on the interdisciplinary panel. "Thanks to continuing research in comparative psychology, genetics, neuroimaging, and paleoanthropology, we know plenty about the evolution of human cognition," said anthropologist Dean Falk of Florida State University in Tallahassee. In rebuttal to Lewontin's complaints about a meager fossil record and the dangers of inferring cognitive capacity from indicators such as skull size, they cited research that they say provides important—if indirect—insights into uniquely human mental capacities.

Geneticist Christopher Walsh of Harvard Medical School thinks some answers to the mystery of human cognition lie in genes that govern brain development in modern human beings. Ongoing work in his and other laboratories has shown that mutations in a number of these genes lead to microcephaly, characterized by a very small brain and mild to severe retardation, and other brain malformations. Tests for "signs of selection" have shown that some of these genes were targets of natural selection during human evolution.

Walsh described recent research suggesting that a well-studied gene called *ASPM* controls the fate of embryonic nerve cells in the multilayered cerebral cortex. The cells must divide in the correct orientation and at the right time as each cortical layer takes form. Selection for changes in *ASPM*, Walsh said, "may provide an evolutionary mechanism that can enlarge the cortex" and help explain the dramatic expansion of hominid brains that began about 2 million years ago.

Anthropologist Leslie Aiello, president of the Wenner-Gren Foundation for Anthropological Research in New York City, agreed that current research can "get us beyond the paleofantasy that Richard Lewontin is talking about." Aiello argued that the fossil and archaeological records are strong enough to show several "major phases in human evolution," including the split between the chimpanzee and the human lines about 6 million years ago and the invention of stone tools beginning about 2.5 million years ago. Moreover, Aiello said, sophisticated reconstructions of ancient climates have matched evolutionary events with environmental changes. "Our evolution has played out against some of the largest climatic changes in the Earth's history," she said, including a major shift to drier and more variable conditions in Africa right about the time that the first tools appeared.

Harvard psychologist Marc Hauser underscored the cognitive gap between humans and other "smart species" such as chimps, elephants, and dolphins—a gap that he described as being "greater than that between those animals and worms." Recent findings in his own lab and others, Hauser said, show that nonhuman animals can solve specific problems in often sophisticated ways (for example, the nectar-mapping dances of honeybees and the ability of some bird species to hide food and retrieve it much later), but they cannot apply those talents to other situations. In contrast to such "laser-beam intelligence," Hauser said, humans have evolved "floodlight intelligence" capable of adapting one solution to many new problems. Even tool use by animals—such as chimpanzees using sticks to fish for termites—is "whoppingly different" from what humans do, Hauser insisted. He hopes that the manifold human differences summarized in his "humaniqueness hypothesis" will yield clues about how our species evolved.

–MICHAEL BALTER

Tracking and Tackling Deprivation's Toll

Children raised in poverty generally do worse in school and in their careers than do children from wealthier backgrounds. But why exactly does poverty affect them that way, and what can be done to break the cycle?

At the meeting, a University of Pennsylvania (Penn) psychologist described how the lack of environmental stimulation and parental nurturing in poor households might adversely affect language development and memory. Another researcher from the University of Oregon (UO), Eugene, explained how bolstering the child-rearing skills of low-income parents might help improve their

CREDIT: DLILL/CORBIS

children's mental abilities. "The more we learn about the pathways through which poverty affects kids, the more effective our interventions can be," says psychologist Jeanne Brooks-Gunn of Columbia University, who was not involved in either study.

Penn researchers have been following 110 African-American children born to mothers on welfare. Approximately half the mothers used cocaine and other drugs while pregnant. Visiting the children at home for an hour each at ages 4 and 8, the researchers rated the households on environmental stimulation (such as the child's access to books and musical instruments) and parental nurturance (emotional care such as praise). Between 2001 and 2004, when the children were between the ages of 11 and 13, the group gave them cognitive tests.

The researchers found that prenatal substance abuse had a negligible impact on either language or memory. But there was a strong positive correlation between environmental stimulation and language ability—that is, children raised in more stimulating environments did better at language tasks. More surprising, children who received better nurturing scored higher on memory tasks. Environmental stimulation seemed to have little or no effect on memory, and parental nurturance seemed to have no effect on the children's language abilities. "Our results show that poverty affects different neurocognitive systems in different ways," says Penn psychologist Martha Farah, lead author and presenter of the study, which is in press at *Developmental Science*.

In a more recent study not yet published, Farah and her colleagues analyzed magnetic resonance imaging scans of 47 children from the group when they were between the ages of 12 and 15. Children who had received less parental nurturing tended to have a larger hippocampus. Farah notes animal studies showing that rat pups that receive less grooming and licking from their mothers develop abnormal hippocampi and poor memory later in life. She speculates that babies who don't get enough nurturing and emotional warmth have difficulty coping with stress, which hurts brain development.

Hoping to blunt the impact of poverty, Jessica Fanning, a doctoral student at UO developed a program to teach parents ways to stimulate and nurture their toddlers. For 2 hours a week over 2 months, Fanning and colleagues taught 14 low-income parents enrolled in a Head Start program to reinforce positive behaviors and accomplishments in their children with specific praise, maintain consistent discipline at home, and use language in creative ways such as responding to something the child might say—such as, "Here goes the ship"—with a more descriptive sentence such as "The ship goes fast."

In tests conducted within a month after the course, the researchers found that the children whose parents had received the training showed gains on attentional and memory tasks; children of a control group did not. "The changes in parenting seem to have had a positive trickle-down effect on the children, at least in the short term," says psychologist Courtney Stevens, who worked on the study and presented the findings at the meeting.

–YUDHIJIT BHATTACHARJEE

CREDIT: DAVID MONNIAUX

OCEAN CO_2 STUDIES LOOK BEYOND CORAL

One million tons of atmospheric carbon dioxide (CO_2) are dissolved into the oceans every hour, a process that helps maintain the Earth's delicate carbon balance. But CO_2 also makes seawater more acidic, and too much of it can wreak havoc on a marine species. Three sessions at the meeting described how marine scientists are trying to assess the effects of acidification.

The ocean's average pH worldwide, now roughly 8.4, has dropped about 0.1 since preindustrial times. Scientists estimate that it could fall another 0.4 by 2100 if carbon emissions continue on their current trajectories. That could put nearly two-thirds of known cold water corals into corrosive waters, Ulf Riebesell of the Leibniz Institute of Marine Sciences in Germany told one colloquium. But although the risks to corals are well-known (*Science*, 4 May 2007, p. 678), the effects on other marine life are just beginning to be characterized.

Gretchen Hofmann of the University of California (UC), Santa Barbara, reported that a one-two punch of lower pH and higher temperature can be fatal for the purple sea urchin, *Strongylocentrotus purpuratus*. Hofmann's lab studied urchins in tanks of seawater at normal pH and at the stronger acidity expected by 2100 under two possible atmospheres described by the Intergovernmental Panel on Climate Change. As the pH fell from 8.1 to 7.8, sea urchin larvae struggled to build their skeletons. DNA microarrays showed that genes involved in biomineralization raised their activity threefold. "The larva is desperately trying to make its body," Hofmann said. Unpublished results from the lab showed that larvae in the most acidic water grew "short and stumpy" skeletons. If the deformities carry over to adults, they could affect the valuable fishery for urchins, which are harvested for their eggs.

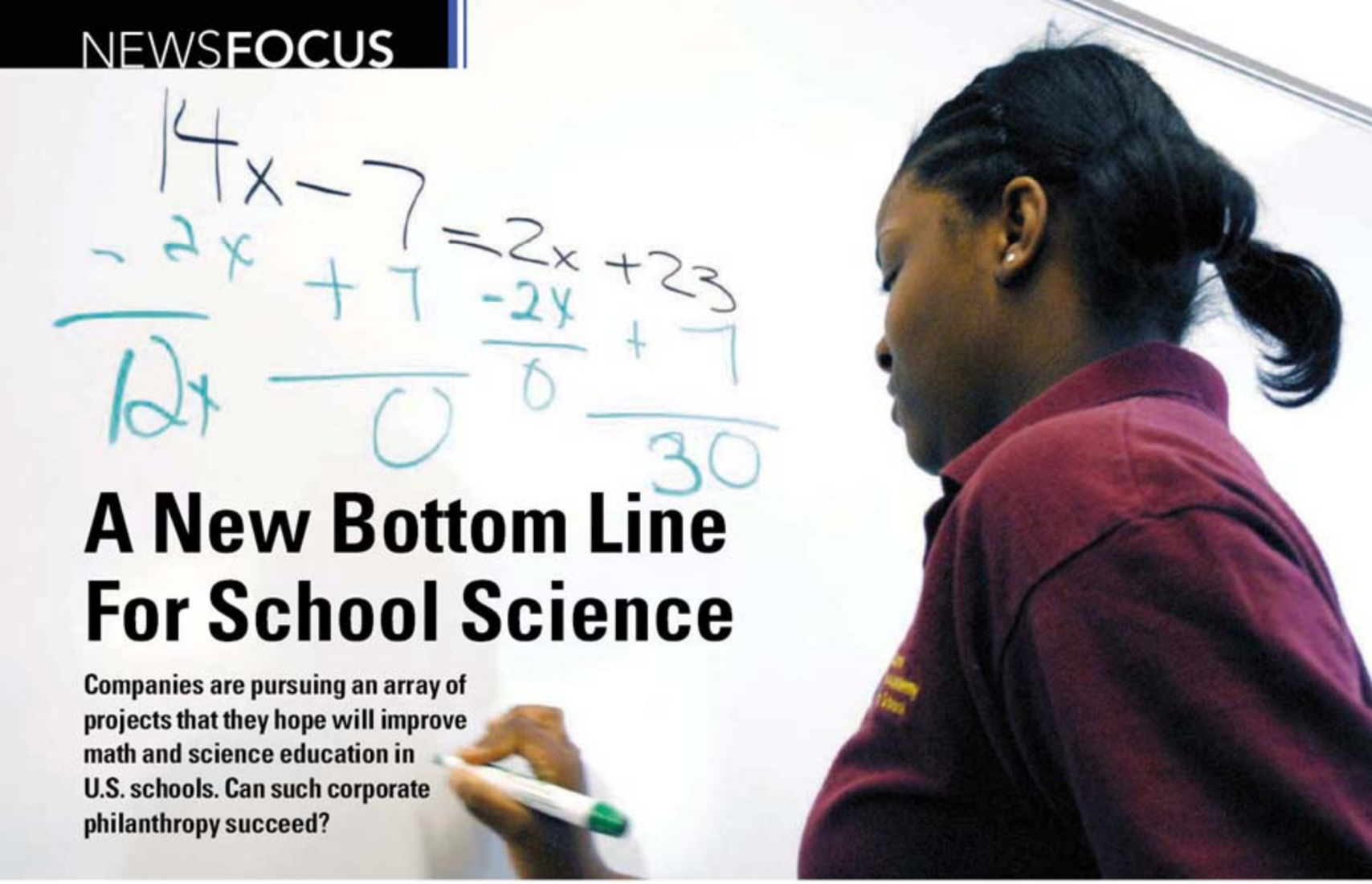
Threatened. Purple sea urchins are among many marine organisms likely to suffer as oceans acidify.

When Hofmann and colleagues warmed the acidified waters, mortality among the larvae skyrocketed. "Gretchen has the story dead on with the urchins," comments Andrew Baker of the University of Miami in Florida, who is studying the effects of temperature and acidity on corals. "Clearly, the effects are worse together than separate."

Hofmann and Victoria Fabry of California State University, San Marcos, are now studying how acidity and temperature affect the pteropod *Limacina helicina*, a peppercorn-sized swimming snail that forms a key part of the food web in the Southern Ocean. In their evolutionary history, Riebesell says, many species of pteropods "have never seen an ocean as acidic as the one they're going to see in the next 100 years."

Paleoclimate researchers are also beginning to study how high CO_2 levels might have impacted species in ancient seas. A team led by James Zachos of UC Santa Cruz is focusing on a 150,000-year period, 55 million years ago, when the amount of carbon released into the atmosphere—nearly 4 gigatons—is similar to the pulse researchers expect from current human emissions. Estimating ocean pH for this extreme event is tricky, Zachos said, because most of the standard indicators—calcium carbonate shells in sea-bottom cores—dissolved away. But he hopes computer modeling and isotope analysis of other shell samples will give his team a handle on the past—and possibly on our torrid future.

–ELI KINTISCH AND ERIK STOKSTAD



A New Bottom Line For School Science

Companies are pursuing an array of projects that they hope will improve math and science education in U.S. schools. Can such corporate philanthropy succeed?

CHICAGO, ILLINOIS—A new math and science charter high school opened this fall on this city's west side. It's named for Exelon CEO John Rowe, who gave $2 million, and Frank Clark, chair and CEO of ComEd, a major Exelon subsidiary that lights the city, who chipped in $200,000. Their gifts, plus $2 million from the Chicago-based utility itself, created the Rowe-Clark Math and Science Academy, which currently shares space with a community health clinic in a squat, two-story building on a tired street in a mostly African-American community. "Math and science is what we do, and we hope to interest kids in careers that will eventually lead them to Exelon," explains Peggy Davis, a lawyer who oversees Exelon's philanthropic efforts as its vice president for diversity.

Big companies have long been involved in helping local schools. But Rowe went a step further by not only specifying the focus of the new school but also selecting the organization, the Noble Network of Charter Schools, to run it. It's a fresh challenge, admits Michael Milkie, a former Chicago math teacher who started the first school in 1999. Noble runs several college-preparatory schools, but none is focused on the sciences.

Much of the burden of devising and implementing a new curriculum has fallen on Vanessa Galarza, the school's math and science department chair and sole science teacher for the 145 ninth graders who constitute the school's inaugural class. "Most of the students have never done a lab report or taken data from an experiment," says Galarza, a former doctoral student in astronomy at New Mexico State University in Las Cruces. And although she is grateful for the well-equipped labs that the Exelon gift made possible, Galarza knows they won't make up for the impoverished academic backgrounds of most of her students. "They've never kept a lab notebook. And I had to teach the X and Y axes," she says. "Also, I don't use a textbook or assign written homework because so many of them wouldn't be able to read it."

Davis, a member of the Chicago Board of Education and a former chief of staff to the current superintendent, understands how far the students at Rowe-Clark need to travel academically before they will be capable of landing a technical job at the utility company. "Ideally," she says, "we'd have loved to do a boarding school starting in first grade. That would have leveled the playing field by letting us deal with all the issues that students bring from home. But we have to be realistic about what we can afford."

Exelon's paternalistic attitude toward the new school is characteristic of the latest wave of corporate philanthropy aimed at improving precollege STEM (science, technology, engineering, and mathematics) education in the United States. Corporate philanthropy is a $14 billion enterprise, and education is the biggest recipient of that largess. Although no accurate figures are available, a sizable slice of that pie is devoted to pre-K–12 (prekindergarten through high school) activities in STEM education.

For decades, most companies supporting STEM education were content to stay in the background. The Intel (formerly Westinghouse) Science Talent Search that awards college scholarships to high school students may be the most prestigious example of such corporate altruism. Other companies have given grants to nonprofit organizations devoted to improving the skills of teachers or strengthening the math and science curriculum.

That's still happening. But Exelon and a growing list of companies without any such track record have put themselves and their employees on the front lines in response to the disappointing performance of U.S. students on more than a decade of international math and science tests. "We believe that American businesses must be active, engaged leaders in this work," proclaims a report issued last summer

CREDIT: HARP INTERACTIVE

A multiplier effect. Exelon's support helped launch the new Rowe-Clark Math and Science Academy, a public charter school on Chicago's west side.

by the Business-Higher Education Forum.

Ernst and Young's Tony Anderson couldn't agree more. "We bring discipline to the issue and accountability. That's what seems to be needed," says Anderson, head of the Chicago office and vice chair of the New York City–based professional services firm, which operates an active mentoring program at several Chicago charter schools.

Clark, an African-American who grew up in a poor, single-family home on the city's South Side and who began at ComEd as a mail clerk, feels a personal obligation to lend a hand. "These kids are at risk, and without help they will struggle," he says, noting that he recognizes himself in some of the students he has met. "I won't be content if I don't do everything possible to give them a shot at getting to the top, like I did."

But is caring and cash enough to make a difference? Experts in corporate philanthropy and educational assessment say that although every little bit helps, many companies may have unrealistic hopes for what their dollars can accomplish. "If you lack firm goals, then measuring whether you've succeeded is very difficult. That's probably the most systemic problem we see across corporate philanthropy," says Gregory Hills of FSG Social Impact Advisors in Boston, which was hired by Ernst & Young to produce a report last year titled *Best in Class: How Top Corporations Can Help Transform Public Education*. Without clear-cut targets and a rigorous way to keep score, Hills warns, "the most that a company is going to achieve is some nice headlines, better ties with local leaders, and some new customers."

Strategic philanthropy

There are three reasons companies invest in STEM education, says Hills. The first, and most common, is what he calls "communal obligation." By spending money in locations where they have plants and offices, companies hope to demonstrate that they are good corporate citizens. But because such philanthropy lacks any specific goals, he says, it's nearly impossible to assess the impact of those dollars on students.

PNC Financial Services found itself in that position a few years into a 10-year, $100 million initiative begun in 2004 called Grow Up Great. After two rounds of grants to Head Start programs in nine states that the bank serves, says Eva Blum, president of the PNC Foundation, company officials realized that the portfolio was so diverse that it would be hard to scale

From an Idea to a School

Not far from the neighborhood where Exelon has helped launch a school focusing on science and math (see main text), another giant Chicago-based corporation is funding a charter school with similar aims but using a different approach. Instead of the donor, Motorola, calling the shots, the new school for grades 6 through 12 is expected to be a test bed for university-based work on delivering STEM education in an urban setting.

Before the charter organization, Perspectives Charter Schools (PCS), even approached Motorola, it first teamed up with science educators Norman and Judith Lederman of the nearby Illinois Institute of Technology (IIT). The Ledermans, who are already working with several Chicago public schools, have spent years developing a curriculum that uses scientific inquiry as the driving principle for teaching every subject. The Ledermans found PCS, which runs four charter schools, through one of its science teachers-turned-administrators, Mary Cummane, who is working on her doctoral degree at IIT. Together, they drew up a proposal for a school that would incorporate their ideas on curriculum and professional development for STEM teachers. And Motorola, which had already promised to support a fund that aims to create 100 public charter schools in Chicago, loved it.

"Sometimes the most important thing a company can do [is] say, 'Yes, we believe all kids deserve a quality education.' "

—EILEEN SWEENEY, MOTOROLA FOUNDATION

It didn't hurt that the director of the Motorola Foundation, Eileen Sweeney, had, a decade ago, helped start another inner-city charter high school, North Lawndale College Preparatory, while running the foundation for Chicago-based United Airlines. That experience taught her the value of assembling and retaining a top-notch teaching staff. She says PCS's ability to do that in the four other schools it runs, combined with the resources that IIT will contribute, makes the new academy "a dream come true" for Motorola, which in November made public its $500,000 donation.

Motorola also hopes the school will augment ongoing efforts to increase student participation, especially by girls and minorities, in technology-related fields. Last fall, Motorola announced it would hand out $3.5 million to 106 projects with that goal as part of its new Innovation Generation Grants program. A call for a second round of proposals, with $4 million to be awarded, went out in December.

Asked what Motorola hopes to accomplish after 5 years by funding the new academy, Sweeney answers, "I'm glad that you said 5 years, because any change takes a while to show up." And whatever improvements occur, she adds, won't be something for which Motorola can claim credit. "We won't be breathing down their necks," she promises. "Sometimes the most important thing a company can do is to put its hand up and say, 'Yes, we believe all kids deserve a quality education.' " **–J.D.M.**

Brain trust. From left, IIT's Judith and Norman Lederman plan a new Chicago charter school to be led by Mary Cummane.

CREDIT: REBEKAH RALEIGH

Building momentum. Science teacher Vanessa Galarza helps Rowe-Clark students set up a lab in their introductory physics class.

up any projects found to be effective. So last year, the initiative, which has several components aimed at getting children ready for school, was refocused on giving Head Start teachers the math and science skills they lack to prepare children for a lifetime of learning.

Thanks to a 3-year, $135,000 PNC grant, the 54 teachers at the Council of Three Rivers American Indian Center's Head Start program in Pittsburgh, Pennsylvania, "have become a lot more comfortable with science," says center director Maggie Gombas. In addition to receiving a flood of science materials, she says, the teachers learned at workshops led by the city's Carnegie Science Center "how to look beyond the obvious" by, for example, turning the tale of the Three Little Pigs into a lesson on wind as a force of nature.

The second type of philanthropy, which Hills calls "brand imaging," is meant "to influence its customers and the external world: government regulators, local officials, and so on." That approach may help companies improve their bottom lines and deflect unfavorable publicity, Hills says, but it makes them less likely to join forces with others and, thus, limits the impact of their philanthropy. "If PR and one-upmanship is the ethos" in the company, says Hills, it leads to "these silo programs."

S. Anders Hedberg, head of corporate philanthropy at Bristol-Myers Squibb (BMS), has tackled the branding problem head-on with a new program for high school students. Since the early 1990s, the New Jersey–based pharmaceutical giant has given out tens of millions of dollars to improve the curriculum and quality of teaching at middle schools. Those programs, directed by the nonprofit National Science Resources Center (NSRC), a joint effort of the Smithsonian Institution and the U.S. National Academies, have been lauded for their progress in addressing two factors—uninspired instruction and poorly trained teachers—that give many preteens another reason to lose interest in math and science. But a few years ago, Hedberg, a cardiovascular pharmacologist who joined the company in 1980, decided that BMS needed "to move the needle forward." His team spent a few million dollars putting together a 12-lesson unit, called R_xeSEARCH, that tells a fictitious story of an outbreak of an unknown, highly contagious disease. The unit, now being piloted in the mid-Atlantic region, combines what the company knows best—drug discovery—with what it has learned about reforming STEM education.

One novel twist in the R_xeSEARCH project is that it's run by a consortium of pharmaceutical and biotech companies, each of which has anted up for the material and for a summer institute to train teachers. The idea is for each partner company to work with school districts in the communities it serves, modifying the curriculum as needed with examples from its own labs. "It's always hard for a company to say to a competitor, 'I have a good idea. Do you want to join me?' So we decided to give it away before it became branded with the BMS name," Hedberg explains. The group has asked NSRC to help implement the project in an arrangement that is still being hashed out.

North Carolina–based Wachovia bank, third largest in the country, is also eschewing the "brand imaging" approach in an effort to encourage partnerships. In 2004, the bank began making competitive 3-year grants of as much as $750,000 to universities and community organizations for new teacher training and professional development, with a focus on science and math. But it attached two strings. "We told them you can't use Wachovia's name [on any promotional material] because it turns off other potential donors," says Dee Lee, who runs the teaching initiative. "It was a huge change in our corporate philosophy." In addition, she says, the size of the grants tapers off

Money Doesn't Always Talk

Companies are eager to describe their philanthropic efforts to improve math and science education. But they can be less forthcoming about the details. That's what *Science* found when it tried to follow the money trail of some large corporate STEM (science, technology, engineering, and mathematics) initiatives.

Boeing Co. spends $21 million a year on education at all levels, says Joyce Walters, director of global community investing at the company and "subject matter expert" for the foundation's education activities. It supports dozens of programs that have attracted national attention, including the Leadership and Assistance for Science Education Reform project in Washington state, and has recently ramped up a preschool learning initiative. But Susan Birkholtz, a company spokesperson, says, "We do not provide lists of grantees, nor the amount of individual grants."

Houston, Texas–based Shell Oil Co., a subsidiary of the global energy giant, has a half-century track record of supporting precollege, university-based, and informal science education programs aimed at increasing the pool of students pursuing technical careers. In 2004, it shifted its mechanism for giving from the Shell Foundation to corporate offices "so that we could get more personally involved and be more of a partner," says Frazier Wilson, the company's social investment manager. In 2006, it joined with *Weekly Reader* to create a Web site for students and teachers called Energize Your Future with Shell that Wilson says tries "to make math and science fun" while building basic skills.

But again, don't bother to ask for too many details. "It is Shell policy not to disclose project budgets," explains company spokesperson Darci Sinclair. Also under wraps is a recent outside evaluation of its philanthropic efforts; Wilson would only say that it had led Shell to become "more engaged and increase its volunteerism." **–J.D.M.**

CREDITS (TOP TO BOTTOM): HARP INTERACTIVE; ERIK RANK/GETTY IMAGES

over the life of the project. "It puts the onus on them to find other partners and not use us as a crutch," she explains.

Hills's third type of corporate charity is the one he says works best. He calls it "strategic philanthropy," and it's distinguished by companies that ask: "What are our strategic business interests, and how does education and the work force fit into that strategy?" As a result, Hills says, "they pick things that they really care about and that they are good at."

That's what The Boeing Company has tried to do, says Joyce Walters, director of global community investing at the aerospace giant, which recently moved its corporate headquarters from Seattle, Washington, to Chicago. "We take a business approach to education investments" that include a Washington state science initiative, districtwide math reform in both cities, and a national program to train principals in urban school systems, she says. And rather than waiting to see whether a particular project succeeds or fails, Walters says Boeing stays involved after the money is handed out. "If one of our suppliers had a problem, we wouldn't just walk away. We'd try to figure out how to solve it. And that might include more resources."

Strategic philanthropy isn't the answer for every company, says Hills. "Some corporations don't believe that their business and social interests can overlap," he explains. "They see it as too self-serving." The Merck Institute for Science Education (MISE), created in 1993 to improve science and math in U.S. elementary and secondary schools, may be a case in point.

Then-CEO P. Roy Vagelos decided that a company so dependent on scientific talent "should be looking much earlier in the pipeline" than the small undergraduate program to attract minorities into the pharmaceutical industry that the New Jersey–based pharmaceutical company had been funding for years. Instead of making MISE an arm of the company's corporate philanthropy, however, Vagelos decided that the new, freestanding institute "should be clearly independent from the business part of Merck." He also recruited a director, Carlo Parravano, a chemistry professor at the State University of New York, Purchase, who had considerable experience in precollege science education.

That hands-off approach seems to have worked well. During the past 15 years, MISE has funded $50 million worth of projects, most aimed at improving the quality of middle school science and math teachers. The institute has leveraged the support of its corporate benefactor to win two multimillion-dollar teacher-training grants from the U.S. National Science Foundation, the gold standard for work in the field of STEM education. And this year, Merck asked MISE to manage all of its science education activities.

Return on investment

Whatever a corporation decides to do, it will want to know at some point what its dollars have accomplished. But evaluation can be a casualty when corporate largess and the U.S. education system collide. "Measurement is a huge problem," says Mary Wright Benner, who runs the education arm of the Conference Board, a coalition of 2000 companies and organizations that has been a major voice in promoting STEM education. "Most schools are focused on what the state or federal government wants them to measure. Funders want to see results, and they may consider evaluation to be part of overhead."

Showtime. Employees from Ernst & Young's Chicago office use the PBS show *Cyberchase* to help Perspectives-Calumet Middle School students with math.

One problem with evaluation is its price tag, says W. Steven Barnett, director of the National Institute for Early Education Research at Rutgers University in New Brunswick, New Jersey, and an adviser to the Grow Up Great program. "To do a real scientific evaluation, you can spend more than what the initiative itself costs," he says. "I think it's better for a company to advise policymakers to spend more on STEM rather than actually try to do it themselves."

Blum, of the PNC Foundation, admits that assessing what Grow Up Great has accomplished hasn't gone as smoothly as she had hoped. "Evaluating children at this age is difficult. And it's hard to isolate the impact of separate interventions," she says. The foundation has received a federal grant to hire an outside evaluator for its math programs, and it is paying for a separate evaluation of its science efforts. The results of these evaluations, she says, will determine the future of the initiative.

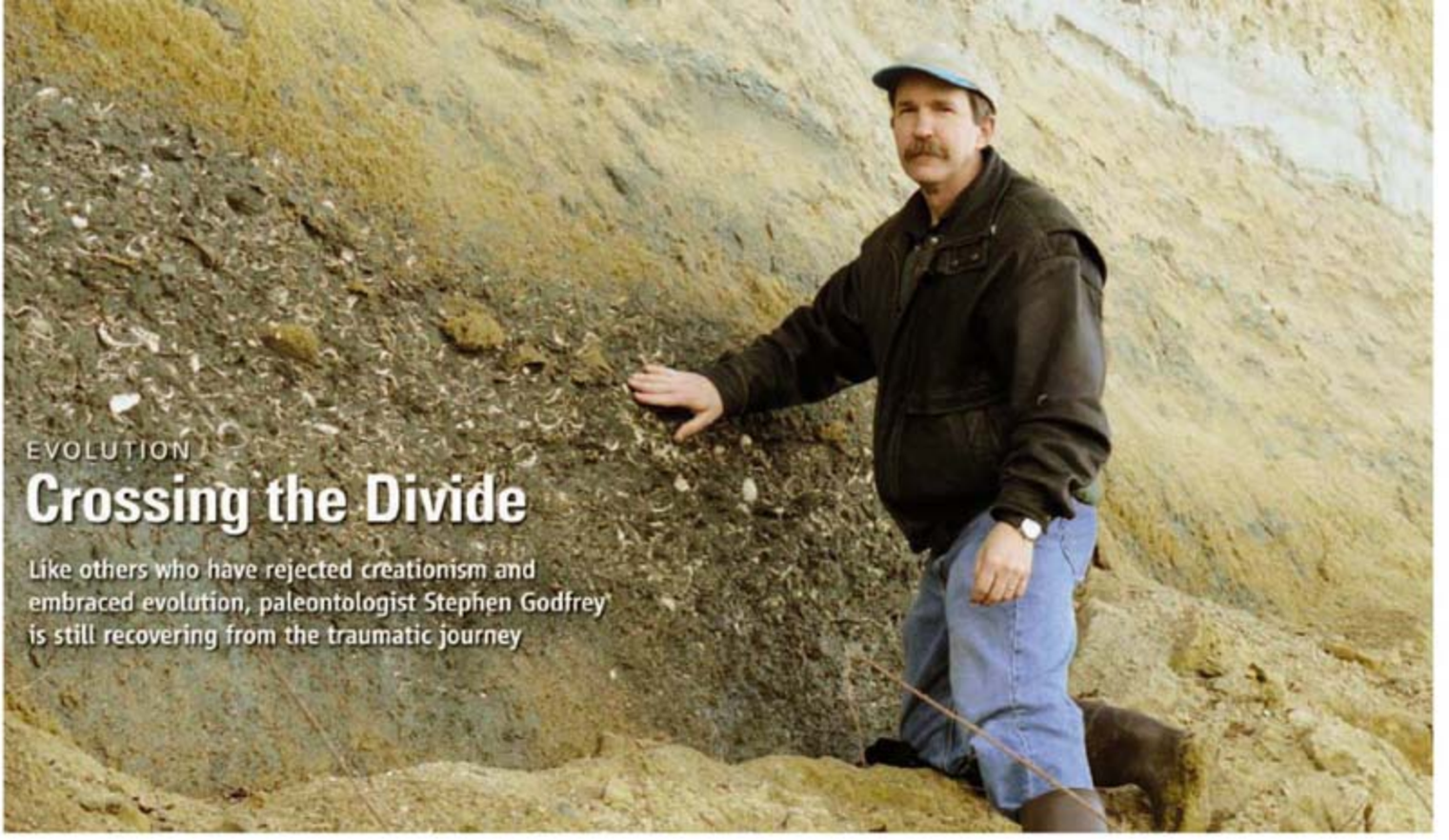
Three Rivers's Gombas isn't sure that her center will even be able to teach other Head Start programs what it has learned. "I don't think it will work out that way," she says. In addition to the staffing problems created by sending teachers out periodically to work with programs at other locations, she says that the training "will lose a lot" of its value if the Carnegie center is no longer involved. Gombas also notes that her teachers feel "saturated with science" and eager to shore up their skills in other content areas.

Wachovia has also hired outside evaluators, Lee says, and initial results suggest that teachers who participated in the program feel more confident about their skills in many areas. But Lee says the initiative is really aiming at systemic change of STEM education. "We want to scale up what they've learned about leadership, decision-making, and working in teams," she says. "And the more outside groups that get involved, the better it will be for the kids."

After spending nearly 2 decades with one foot in industry and the other in education, BMS's Hedberg knows that systemic change is incredibly hard to achieve. And he offers a few lessons from his dual life to any company thinking of engaging in STEM philanthropy.

One is to start with what you know. "This curriculum is based on our industry's genetic code: how we develop a new drug," he says. "We're not telling teachers what to do. And it's not a marketing program. We're just giving them a version of the science, which is not suitable for the classroom in its raw form, in a format that they can use." Another is that one-shot approaches don't work. "We're hoping to follow the NSRC model, which incorporates professional development, curriculum, materials, community involvement, and assessment," he explains. "None is enough by itself."

A third lesson is that reformers need patience to overcome the tremendous inertia within both the education and the pharmaceutical sectors. "This is going to take years," Hedberg admits. "But I think the idea of multiple companies with equal voices is pretty darn new for the industry." **–JEFFREY MERVIS**



EVOLUTION

Crossing the Divide

Like others who have rejected creationism and embraced evolution, paleontologist Stephen Godfrey is still recovering from the traumatic journey

SOLOMONS, MARYLAND—On a clear January day, Stephen Godfrey is dressed for fossil-hunting: frayed baggy jeans, a puffy green vest, and a leather jacket that's seen better times. A paleontologist and curator at the modest Calvert Marine Museum here, Godfrey frequents the nearby Calvert Cliffs, which rise from the shoreline of Chesapeake Bay and hold everything from ancient shark teeth to dolphin skulls. "You start collecting them because, well, they're beautiful," he says of his beloved fossils.

It was the study of fossils that, 25 years ago, set Godfrey on an anguished path. Raised in a fundamentalist Christian family in Quebec, Canada, embracing a 6000-year-old Earth where Noah's flood laid down every fossil, Godfrey began probing the underpinnings of creationism in graduate school. The inconsistencies he found led step by step, over many years, to a staunch acceptance of evolution. With this shift came rejection from his religious community, estrangement from his parents, and, perhaps most difficult of all, a crisis of faith that endures.

Powerful emotions bind together young-Earth creationists, members of a movement making inroads from Kenya to Kentucky, where a $27 million Creation Museum opened last year. Scientists and educators have responded mainly by boosting biology's place in the classroom and building rational arguments for evolution. But reason alone is rarely enough to sway believers. That's because letting go of creationism carries enormous emotional risks, including a loss of identity and community and an agonizing, if illusory, choice: science or faith.

People like Godfrey tend not to advertise their painful transition from creationist to evolutionist, certainly not to scientific peers. When doubts about creationism begin to nag, they have no one to turn to: not Christians in their community, who espouse a literal reading of the Bible and equate rejecting creationism with rejecting God, and not scientists, who often dismiss creationists as ignorant or lunatic.

"Nothing else I have done in my life has made me such an outsider," says Brian Alters, director of McGill University's Evolution Education Research Centre in Montreal. Alters has written books on teaching evolution and testified in the 2005 Dover, Pennsylvania, trial against bringing intelligent design—a form of creationism—into the classroom. But few of his friends or his enemies know that Alters, who had a fundamentalist Christian upbringing in southern California, rejected creationism in college. More than 2 decades later, he says, "I still have childhood friends and relatives who won't speak to me."

Turning point. These fossilized footprints from an early reptile sent Godfrey into a tailspin and forced a letting go of creationism.

Faithful upbringing

Religion anchored Godfrey's childhood. He was the third of five children—"a great place to be overlooked," he jokes. Every evening after dinner, his father, a Sunday school teacher, pulled out the Bible. "We would go systematically through two readings of books," says Godfrey, and devote time to prayer. The family attended church twice on Sundays, in the morning and in the evening, and one parent or the other often dropped in on a Bible study class midweek.

From a young age, Godfrey had a keen interest in biology. He adored touring natural history museums and collected pinecones, rocks, minerals, and anything else he could find outdoors. Skeletons in particular captivated him for their visual aesthetic. During visits to his mother's family in New York state, he began gathering the skeletal remains of groundhogs and squirrels left by the side of the road, carefully wrapping them in black garbage bags for the trip home to Quebec.

His parents saw no conflict between their son's love of biology and their beliefs and encouraged his interests. "I guess they figured that the young-Earth creationist position was strong enough, was robust enough, that he would believe in young-Earth creationism and he would be a biologist, and that would

CREDITS (TOP TO BOTTOM): J. COUZIN/SCIENCE; COURTESY OF STEPHEN GODFREY

Digging for answers. Paleontologist Stephen Godfrey rests his hand on shells embedded in Maryland's Calvert Cliffs, which were underwater 14 million years ago.

be fine," says Godfrey.

Now 48, Godfrey came of age after young-Earth creationism took hold in North America in the early 1960s. Its leaders argued that during the previous 150 years, Bible-believing Christians had gone too far in accommodating science in their interpretation of scripture and pushed for a literal reading of the Bible, says Ronald Numbers, a historian of science at the University of Wisconsin, Madison. Fossils, for example, are the remains of plants and animals left out of Noah's ark. The description of Adam and Eve in Genesis suggests that humans had never been subject to evolution. Using calculations drawn from genealogy, young-Earth creationists consider the planet to be 6000 to 10,000 years old. (Geologists say it is about 4.5 billion years old.)

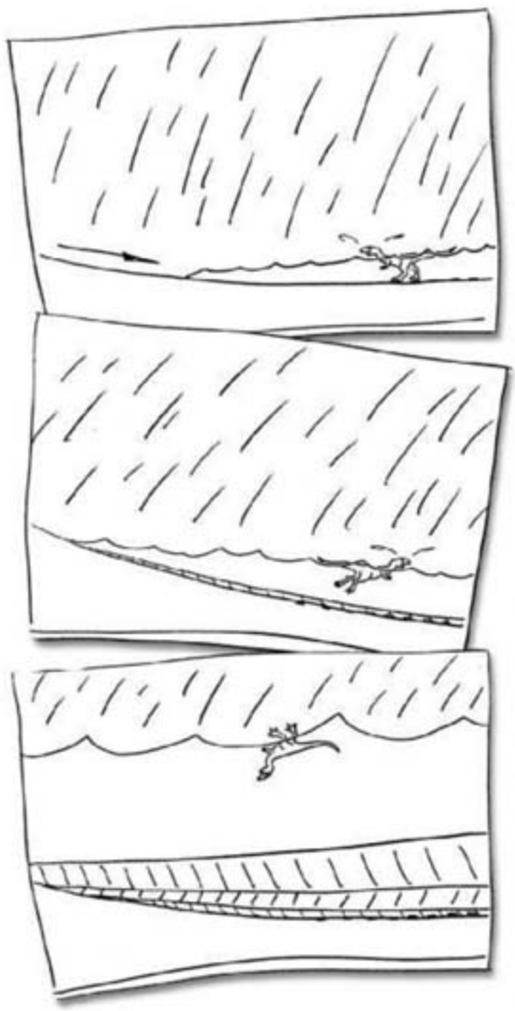
Godfrey, who subscribed wholeheartedly to these views, vividly recalls his earliest encounter with evolution. In the first grade, when he was about 6 years old, a student teacher said that apes were the ancestors of people. "I remember having this visceral reaction … and saying, 'No, that can't be.'" Around the dinner table that night, his family discussed the experience, concluding that the teacher must have been mistaken. "It couldn't be true because apes aren't evolving into humans today; they're apes," Godfrey remembers. And that was that.

Although creationism might seem bizarre to individuals who have never believed in it, for those who do, its power is almost beyond words. Alters remembers, as a young teenager, sitting in on a sermon by Robert Schuller, a televangelist whose California church is fairly liberal. Listening to Schuller endorse the views of scientists who consider rocks to be millions of years old, Alters began to cry, horrified that the preacher would lie. "It was almost as if he stood there and said Jesus Christ didn't exist," he recalls. For biblical literalists, belief is generally an all-or-nothing proposition.

Identity crisis

Godfrey entered college convinced that scientists were engaged in a vast conspiracy to promote evolution. At Bishop's University in Sherbrooke, Quebec, he majored in biology and lived at home, several kilometers away. In one sense, his studies had little effect on his faith. "You can learn facts, and you can do really well on exams and not believe" what you're learning, he says. But then, his classes also raised niggling questions that biblical literalism could not easily answer.

For example, there was the quandary of death. A literal reading of Genesis indicates that no animals perished before Adam and Eve ate the fateful apple—in other words, that there were no carnivores preying on other animals. But in his biology classes, Godfrey learned of predators perfectly framed to kill: cats with stereoscopic vision, enlarged canines, and claws; spiders that weave webs as traps; and sharks that replace serrated teeth throughout their life. "They're not eating seaweed," says Godfrey, who puzzled over how these animals had emerged if God hadn't intended them to prey on others. "That was the first thing at university that really started to disturb me," he says.



Rebutting Noah's flood. Godfrey drew these sketches to show how Noah's flood cannot explain fossil footprints, as they're found in different layers of rock and depend on an animal resting its weight on the ground.

In his final year, Godfrey gave a presentation on the origin of flight, arguing that *Archaeopteryx*, the earliest known bird, could not have evolved from the dinosaurs. Although impressed by similarities between *Archaeopteryx*'s anatomy and that of dinosaurs, he pushed this to the back of his mind. By this time, Godfrey was at a crossroads and determined to find out for himself whether the claims of biologists and paleontologists were true. He enrolled in graduate school in paleontology at McGill University and was taken in by Robert Carroll. Carroll had heard that Godfrey was a creationist but didn't give it much thought, he says now. In Carroll's lab, Godfrey prepared and described fossils of an ancient amphibian called *Greererpeton*. The fossils "could have come from the moon," says Carroll. Analyzing them out of context had little impact on Godfrey's views.

Then Godfrey's world came crashing down.

His first summer in graduate school, he was invited to join a field expedition in rural Kansas, where University of Toronto paleontologist Robert Reisz and some students were digging for pelycosaurs, 300-million-year-old animals that display some features of mammals that evolved later. Living in tents on a farmer's field in searing heat and humidity and surrounded by cows, the group visited the nearest town, Garnett, weekly for food and other supplies. At night, the sky glowed with stars, and Godfrey pointed out the constellations to his companions.

By day, quarrying through thin layers of rock, "we started to come across footprints of terrestrial animals," says Godfrey. "You can't imagine a global flood and animals finding ground to make footprints on. … That, more than anything, any other experience in my life, really shook me to the core." Godfrey agonized about where these footprints might have come from. Some creationists argue for floating mats of vegetation during the flood, but Godfrey found that unconvincing.

"He was one of the brightest students that I'd ever seen," says Reisz, who at the time knew that Godfrey was a devout Christian but had no idea of the crisis triggered by his fieldwork. "The ease with which he learned, the ease with which he accumulated new ideas, … all spoke to a superior intelligence."

Godfrey held out from embracing evolution, however, until after moving in 1989 to Drumheller, Alberta, dubbed the "dinosaur capital of the world" because of its diversity of fossils. Godfrey often drove southeast to Dinosaur Provincial Park, passing through a landscape of sediments laid atop one another: deposits from freshwater and terrestrial environments in one, marine organisms and mollusks in another, and a third that mimicked the

CREDITS: STEPHEN GODFREY

Earth according to Bible. The expansive Creation Museum in Petersburg, Kentucky, opened last May.

first, a mix of fossils from fresh water and land. "These animals were living here in this same place, but they couldn't have all been there at the same time," he says, a fact that was irreconcilable with flood geology. It was then that "the rest of the young-Earth creationist ideas kind of exploded."

Godfrey ran through bitterness, anger, and disappointment about having been deceived for so many years. He sought out creationists and confronted them. Late in graduate school, he and his devout Christian wife, mother-in-law, and mother attended a weekend symposium at a Bible school in New York state, where Godfrey says he angrily stood up at the end of a talk and argued passionately with the speaker.

It was there, and in conversations during holiday meals, that Godfrey's parents realized that he had changed. Deeply unhappy, they worried whether their son could endorse an old Earth and remain a Christian. Their message was, "It's all or nothing," says Christopher Smith, Godfrey's brother-in-law and a pastor at the University Baptist Church in East Lansing, Michigan. "I do remember a discussion one year at Christmas; the tone quickly turned angry," Smith says. Godfrey's father eventually asked that he stop mentioning evolution, as the topic was too upsetting to the family, who believe that their afterlife depends on embracing creationism.

Parents often cannot cope with such an upheaval in a child. "The day I had to tell my mother I wasn't a young-Earth creationist was the scariest day of my life," says Denis Lamoureux, who teaches science and religion at St. Joseph's College in the University of Alberta in Edmonton, Canada. His mother was so embarrassed by his work in biology that she told her friends her son was still in the profession he once belonged to: dentistry. Some compare these conversations to informing fundamentalist Christian parents that they are gay—but perhaps even more wrenching.

Jagged resolution

Trying to articulate where his religious beliefs stand now, Godfrey's eyes fill with tears. "It's been so long, a lifelong struggle, to sort out," he says. He has flirted with atheism but found it too depressing. Several years ago, he stopped attending church for a year before returning. He believes in God today, he says, but tomorrow may be different.

Complicating matters are the people he most loves and their stance on creationism. Godfrey and his wife met as teenagers in a church youth group. They and their five children have always attended an evangelical, young-Earth creationist church. About 6 months ago, Godfrey seethed through 12 weeks of a DVD presentation on creationism. During an early session, he raised objections in front of a church youth group that included his 15-year-old daughter. The group was not brought back for later showings.

"I was really torn," he says, "because I would have loved to have been given the opportunity to say, 'Okay, I'm now going to do a presentation on the other side.' But they don't want to hear it. It's too threatening and it's too upsetting."

Like many creationists-turned-evolutionists, Godfrey is conflicted about how, and how forcefully, to press his case. In 2005, he and his brother-in-law Smith published *Paradigms on Pilgrimage*, a book describing their own transition and making the case for evolution. His father prayed that it would not be published, and Godfrey did not send his parents a copy. He thought his book would change minds among creationists but isn't sure it has.

"I haven't" read it, says his younger sister Esther Godfrey, of Sherbrooke. "I'm feeling it's a very odd way of viewing the Bible, if you can choose which parts you believe literally and not literally." Esther Godfrey is not sure what turned her brother away from a young Earth, as they've never discussed it. "I know he saw something at some point, maybe a fossil, and thought the Earth has to be old," she says. "That is what I've heard."

Just as he longs for biblical literalists to be more receptive to evolution, Godfrey also wishes that biologists would join the discussion. He was incensed 5 years ago when, participating in an evolution-creationism debate at Bishop's University, where he once argued against the fossil record, no one from the biology department attended.

"I continue to think that scientists have made a serious mistake in not engaging the issue," agrees entomologist Susan Fisher of Ohio State University in Columbus. Fisher, always an evolutionist, was shocked to learn that more than half the students in her 700-person introductory biology class identified themselves as creationists. Last year, she received funding from the John Templeton Foundation to bring in scholars, most of them Christians who reject creationism, to speak to the students. "We need to figure out among students changing their minds, what does that?" says Jason Wiles, who studies evolution education at McGill and Syracuse University in New York state and was himself once a creationist.

But sometimes, former creationists believe, changing minds is not worth the heartache it brings. Godfrey no longer considers evolution worth mentioning to his parents, now 78 and 79 years old, and he asked that they not be contacted for this article. "You can live your life just fine and not know squat about evolution," he says.

When it comes to his children, Godfrey's not sure what they believe nor how firmly to steer them. Certainly, he says, they are exposed to creationist teachings. Of all his children it's his youngest, 4-year-old Victoria, who shows the strongest penchant for science. Wandering the beaches near her home, she often asks to bring home bones she finds, just as her father did years ago. Will her view of the world make room for evolution? Godfrey watches and waits and wonders whether to step in.

–JENNIFER COUZIN

CREDIT: MARK LYONS/GETTY IMAGES

GEOLOGY

Japan's Ocean Drilling Vessel Debuts to Rave Reviews

Early finds from *Chikyu*'s first scientific voyage hint at a coming treasure trove of data on the generation of earthquakes, tsunamis, and other geological phenomena

SHINGU, JAPAN—Two years after leaving the shipyard, Japan's first deep-ocean drilling vessel has finally got down to business. Wrapping up its first scientific voyage earlier this month, the *Chikyu* stopped at Shingu, a port town on the Kii Peninsula, about 400 kilometers southwest of Tokyo, where researchers rhapsodized about findings that promise a better understanding of processes that generate earthquakes and tsunamis. "I've been on five [ocean drilling] cruises before, and these are the most exciting results I've seen so far," says Elizabeth Screaton, a geologist at the University of Florida, Gainesville.

In three back-to-back expeditions over 19 weeks along the Nankai Trough, one of the planet's most active fault zones, the $550 million *Chikyu* bored 33 holes, retrieved cores as deep as 1000 meters below the sea floor, and extracted thousands of samples. Scientists will be analyzing the data and samples for years. But preliminary analyses in *Chikyu*'s labs are already offering insights into fundamental seismic processes.

After receiving the *Chikyu* in July 2005, the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) equipped it and sent it on shakedown cruises. The agency then leased it to an oil-exploration operation to subsidize the scientific work. *Chikyu*, which JAMSTEC runs for the Integrated Ocean Drilling Program (IODP), started scientific drilling last September in the Nankai Trough, where the Philippine Sea Plate is subducting beneath the Eurasian Plate. Sediment on the Philippine Plate gets scraped off and piled into what is called an accretionary prism on the Eurasian Plate.

Ruptures along the boundary and in megasplay faults—which branch off the plate boundary and rise to the sea floor—have triggered some of Japan's most devastating earthquakes, including a 1944 magnitude-7.9 temblor centered about 100 kilometers southeast of Shingu. The shaking and the 9-meter tsunami that resulted from shifts in the accretionary prism killed more than 1200 people, devastated the Kii Peninsula, and caused damage as far as Osaka, 110 kilometers north of Shingu. Scientists drilled into the fault that ruptured in 1944 and the surrounding rocks and sediment to gain a better understanding of the subduction zone's evolution and the mechanisms of earthquake and tsunami generation, says Asahiko Taira, a geologist and director-general of JAMSTEC's Center for Deep Earth Exploration.

Drill sergeants. Geologists Elizabeth Screaton and Gaku Kimura examine cores returned by *Chikyu* (*inset*).

Initial findings have perplexed—and delighted—researchers. Speaking at a December press conference after the second expedition, Siegfried Lallemant, a geologist at the University of Cergy-Pontoise in Paris, France, said he and his colleagues were surprised at the young age of the Nankai Trough's features. Material in the accretionary prism appears to date back only 4 million to 6 million years. "I was expecting much older [material]," Lallemant said. "The prism is growing very fast in this area, much faster than I expected."

The megasplay fault system is also more complex than anticipated. Previous work indicated that the 1944 quake occurred along a single shallow fault that branches off the plate boundary. But core samples revealed a second fault. "We're hoping further analysis will tell which fault slipped in 1944 and what is the relationship between the two faults," says Gaku Kimura, a geologist at the University of Tokyo and co–chief scientist of the third expedition. Screaton, the other co–chief scientist, says that temperatures deep in the boreholes are much cooler than models predicted, suggesting that the current understanding of heat flow and fluid movements in the fault zone is flawed.

The early insights are largely thanks to *Chikyu*'s lab capabilities, particularly a computed tomography (CT) scanner that produces 3D images of the core before its casing is removed. After viewing the images, microbiologists, geochemists, and structural geologists agree on where and how to cut the core to preserve materials and features of interest, says Daniel Curewitz, a structural geologist who served onboard as project manager. "We can make a plan before the [core] is cut open," he says.

CT scanners were too unwieldy and costly for routine use on previous expeditions. "I can't emphasize how much the CT scan has changed how we do science out here," says Screaton. "Now sometimes I wonder if previously there were fault zones that we missed because when we opened the core, [sediment] got disturbed. We were working blind." Curewitz says IODP planners are now considering adding a CT scanner to the *JOIDES Resolution*, the other IODP drill ship, which is being refurbished.

Chikyu's full capabilities have yet to be tested. Its drill is designed to reach 7 kilometers below the ocean floor, but the deepest it has gone so far is just over 1 kilometer. Going deeper might require use of a riser, a tube enclosing the drill pipe that circulates lubricating mud that flushes cuttings from the drilling face and shores up sediments. A new riser is expected to see action in 2009, after the vessel returns from another round of oil exploration. "It's too bad we don't have the ship for science full-time, but there are budgetary issues," says Taira. *Chikyu* will resume scientific drilling along the Nankai Trough in November.

–DENNIS NORMILE

CREDITS (TOP TO BOTTOM): D. NORMILE/SCIENCE; JAMSTEC/IODP

LETTERS

edited by Jennifer Sills

The Legitimacy of Genetic Ancestry Tests

I FOUND THE POLICY FORUM "THE SCIENCE AND BUSINESS OF GENETIC ANCESTRY TESTING" (D. A. Bolnick *et al.*, 19 October 2007, p. 399) to be a shallow discussion of a very complex topic.

Bolnick *et al.* criticize one company in particular—DNAPrint Genomics, Inc.—for promoting the idea that race is rooted in one's DNA (*1*). In fact, DNAPrint Genomics takes great pains to explain to customers that both genetics and race are imperfectly correlated with geography. DNAPrint's Web site (*1*) and print publications (*2*) clearly explain that the bases for human-derived notions of "race" incorporate genetics as well as geography, religion, culture, and even socioeconomics. Bolnick *et al.*, however, deceptively imply that genetics does not constitute even a component of the equation.

Bolnick *et al.* also allege that genomic ancestry panels present a biased picture of nonneutral mutations, which is not the case (*2*, *3*). They then imply that genomic ancestry methods rely on imperfect—i.e., insufficiently large—databases and thus produce misleading results. However, the onus on the database developer is not to build a perfect database, but rather to quantify how imperfect the database is. DNAPrint Genomics has spent enormous resources doing just that (*2*).

CREDIT: N. KEVITIYAGALA/SCIENCE

Likewise, Bolnick *et al.* focus on the fact that alleles are continuously distributed, and that "companies sometimes fail to mention that an allele could have been inherited from a population in which it is less common." This statement is misleading. The continuity of allele frequencies is the reason why, when estimating genomic ancestry with respect to a given population model, we use large numbers of well-characterized markers appropriate for that model, with algorithms capable of accommodating uncertainty. DNAPrint Genomics has gone to great lengths to determine confidence intervals and to quantify the bias and mean square error of our estimates. If a customer's likelihood of ancestry from a certain parental population is very small, the customer's report will reflect this.

I also object to the implication that the detection of lower levels of "Native American" ancestry in Asia and Europe is evidence that the AncestryByDNA test is illegitimate. Native Americans are derived from southwestern Siberia and central Asia, an area of the world that likely also contributed through population expansions and migrations to many other populations (*2*, *3*). Numerous "Native American" Y and mtDNA haplogroups have been found in Europe and Central Asia [figures 4-3 and 4-4 of (*2*) and (*4*, *5*)]. The confusion lies in the choice of the term "Native American." Naming parental populations with descriptors based on modern-day populations might lead to misinterpretations (2), but DNAPrint Genomics makes every effort to explain this complex topic to a lay customer base (*1*).

I am also offended by the implication that scientists who work for companies are corrupt. It is wrong to assume that any science coming from a company is suspect because money is the motivator of its generation, while overlooking the possibility that an academic scientist may falsify data in order to procure a grant, ensure tenure, or otherwise enhance stature or financial position. Generalizing about individuals on group membership in this way is the intellectual equivalent of bigotry.

Bolnick *et al.* believe that anyone who says they belong to a group should belong to that group—regardless of whether or not their deep ancestors (as reported by DNA tests) were part of the parental population associated with that group. The irony is that we do not disagree. In some cases, genetic testing is simply not relevant—not because it is flawed, but because it reports only one aspect of "race" or "ethnicity." Genomic ancestry tests demonstrate that admixture is the rule rather than the exception and hence support that idea that human-derived notions of "race" are based on the subjective and ever-changing concepts of social and political identity.

TONY FRUDAKIS

Chief Scientific Officer, DNAPrint Genomics, Inc., Sarasota, FL 34243, USA.

References

1. DNAPrint Genomics, Frequently Asked Questions (www.ancestrybydna.com/welcome/faq/).
2. T. Frudakis, *Molecular Photofitting: Predicting Ancestry and Phenotype from DNA* (Academic Press, Burlington, MA, 2007).
3. M. Jobling, M. Hurles, C. Tyler-Smith, *Human Evolutionary Genetics: Origins, Peoples, and Disease* (Garland Publishing, New York, 2004).
4. M. Reidla *et al.*, *Am. J. Hum. Genet.* **73**, 1178 (2003).
5. P. Underhill *et al.*, *Ann. Hum. Genet.* **65**, 43 (2001).

Response

OUR PURPOSE IN WRITING THIS POLICY Forum (19 October 2007, p. 399) was not to persuade the public that tests for genomic ancestry are wholly illegitimate, as Frudakis assumes. Rather, it was to call attention to this influential commercial enterprise and the need for consumers (and the public) to better understand the capabilities and limitations of the available tests. We also hoped to inspire genetics and anthropological associations to discuss these issues with their members.

We did not suggest, as Frudakis states, that there is no connection between genetics and societal interpretations of race. Racial identity is shaped by a variety of factors, including social relationships, life experiences, and biological ancestry. Although DNAPrint's Web site states that race reflects more than genetics, it still (as of 2008) leads consumers to believe that race is inscribed in one's DNA. DNAPrint defines the "biogeographical ancestry" measured by their test as "the biological or genetic component of race" (*1*), and their underlying model reinforces the archaic racial view that four discrete "parental" populations existed in the past. The assertion that there is some sort of discrete genetic component to race is problematic, and there is no evidence that only four isolated populations existed at any point in the evolutionary history of our species (*2*, *3*). Furthermore, an extensive derivative literature makes it clear that many people think that the AncestryByDNA test identifies their racial makeup (*4*–*7*).

Frudakis asserts that our Policy Forum claimed that genomic ancestry panels are rife with biased, non-neutral mutations. We made no such claim. Instead, we noted that some ancestry informative markers (AIMs) involve loci that have undergone selection (*8*). On the basis of the information provided by DNAPrint Genomics, it is clear that some AIMs are skin pigmentation alleles and others are blood protein alleles involved in malarial resistance (*1*, *4*, *9*–*11*). It is therefore important to consider whether these markers measure ancestry alone, or whether they also reflect shared environmental exposures (and thus are not always indicative of shared ancestry). Because the AncestryByDNA test does not differentiate between different evolutionary reasons for shared alleles, the test results may be misleading.

Frudakis then objects to the suggestion that genetic ancestry tests are problematic because they may yield incomplete results due to limited sampling. He is right that perfect databases will never exist, and we agree that companies should quantify the uncertainty and limitations imposed by their databases. DNAPrint Genomics does calculate the bias, error, and confidence intervals of their estimates, but incomplete geographic sampling creates systematic bias that is difficult to quantify statistically. The sampling of a few, widely dispersed populations for marker selection likely influences the test's results. It is also U.S.-biased because it represents a specifically American racial understanding of human difference. Furthermore, most mitochondrial DNA and Y-chromosome tests do not provide any such statistics, so consumers are often unaware that those tests may yield incomplete or uncertain results.

Frudakis suggests that the problem with the AncestryByDNA test is not that it detects shared ancestry between Native Americans and Eurasians, but that this ancestry is referred to as "Native American." We agree. Because the shared alleles predate the divergence of these populations and likely originated in Central Asia, it is misleading to use them as markers of "Native American" ancestry.

Finally, it is clear that scientists who work for companies are not inherently more easily corruptible than academic scientists. That said, there are differences in how academic and commercial products are evaluated. Peer review may be imperfect, but it does require academics to convince experts that their conclusions are supported by the data. Commercial products are not subject to the same system of peer review, and it can be difficult to evaluate conclusions based on proprietary databases. Consequently, conflicts of interest may lead to different outcomes in academia and the business world.

DEBORAH A. BOLNICK,[1] DUANA FULLWILEY,[2]* JONATHAN MARKS,[3] SUSAN M. REVERBY,[4] JONATHAN KAHN,[5] KIMBERLY TALLBEAR,[6,7] JENNY REARDON,[8] RICHARD S. COOPER,[9] TROY DUSTER,[10,11] JOAN H. FUJIMURA,[12] JAY S. KAUFMAN,[13] ANN MORNING,[10] ALONDRA NELSON,[14] PILAR OSSORIO[15]

[1]Department of Anthropology, University of Texas, Austin, TX 78712, USA. [2]Departments of Anthropology and African and African-American Studies, Harvard University, Cambridge, MA 02138, USA. [3]Department of Anthropology, University of North Carolina, Charlotte, NC 28223, USA. [4]Department of Women's Studies, Wellesley College, Wellesley, MA 02481, USA. [5]Hamline University School of Law, St. Paul, MN 55104, USA. [6]Department of American Indian Studies, Arizona State University, Tempe, AZ 85287, USA. [7]Departments of Environmental Science, Policy and Management, and Rhetoric, University of California, Berkeley, CA 94720, USA. [8]Department of Sociology, University of California, Santa Cruz, CA 95064, USA. [9]Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, IL 60153, USA. [10]Department of Sociology, New York University, New York, NY 10012, USA. [11]Department of Sociology, University of California, Berkeley, CA 94720, USA. [12]Department of Sociology, University of Wisconsin, Madison, WI 53706, USA. [13]Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC 27599, USA. [14]Departments of Sociology and African-American Studies, Yale University, New Haven, CT 06520, USA. [15]University of Wisconsin Law School, Madison, WI 53706, USA.

*To whom correspondence should be addressed. E-mail: dfullwil@fas.harvard.edu

References

1. DNAPrint Genomics, Frequently Asked Questions (www.ancestrybydna.com/welcome/faq/).
2. K. Weiss, M. Fullerton, *Evol. Anthropol.* **14**, 165 (2005).
3. A. Templeton, *Nature* **416**, 45 (2002).
4. K. TallBear, dissertation, University of California at Santa Cruz (2005).
5. N. Wade, *The New York Times*, 1 October 2002, p. F4.
6. A. Harmon, *The New York Times*, 12 April 2006, p. A1.
7. N. Wade, *Before the Dawn: Recovering the Lost History of Our Ancestors* (Penguin Press, New York, 2006).
8. J. Akey *et al.*, *Genome Res.* **12**, 1805 (2002).
9. E. J. Parra *et al.*, *Am. J. Hum. Genet.* **63**, 1839 (1998).
10. E. J. Parra *et al.*, *Am. J. Phys. Anthropol.* **114**, 18 (2001).
11. T. Frudakis *et al.*, *J. Forensic Sci.* **48**, 1 (2003).

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Soil Erosion: A Carbon Sink or Source?

THE REPORT BY K. VAN OOST *ET AL.*, "THE impact of agricultural soil erosion on the global carbon cycle" (26 October 2007, p. 626) raises two serious concerns.

First, the eroded soil is severely depleted of its soil organic matter (SOM) pool (*1*–*3*), which is preferentially removed by surface runoff because it has low density and is concentrated in the surface layer. Along with the SOM pool, severely eroded soil loses its nutrient and water retention capacity. Thus, an eroded soil generally has lower productivity even with additional input of chemical fertilizers and organic manures (*4*, *5*). Because of low productivity and reduced return of root biomass, the rate of replacement of SOM pool is low on severely eroded soils. This is in contrast to the assumptions made by these and other authors (*6*–*10*).

Second, the process of soil erosion by water entails three distinct stages: (i) detachment, (ii) transport/redistribution, and (iii) deposition. The first two stages lead to breakdown of structural aggregates and exposure of the hitherto encapsulated SOM within the aggregate to microbial processes with an attendant increase in mineralization and emission of CO_2 (*3*, *11*–*13*). During the third depositional stage, the rate of aerobic decomposition may be low and reduce the emission of CO_2. However, prevalence of anaerobic conditions accentuates methanogenesis and denitrification leading to efflux of CH_4 and N_2O (*14*), with the relative global warming potential, respectively, of 21 and

310 times that of CO_2. Taking all three stages into consideration, soil erosion is a strong source rather than sink of atmospheric CO_2, and it also exacerbates the problem of nonpoint source pollution and hypoxia in coastal zones. With severe adverse impacts on agronomic production and food security (*15*), especially in sub-Saharan Africa and South Asia (*16*), it is important that the accelerated soil erosion is effectively controlled and eroded soil judiciously restored.

RATTAN LAL[1] AND DAVID PIMENTEL[2]

[1]The Ohio State University, Columbus, OH 43210, USA. [2]Cornell University, Ithaca, NY 14853, USA.

References

1. F. E. Rhoton, D. D. Tyler, *Soil Sci. Soc. Am. J.* **54**, 223 (1990).
2. E. DeJong, R. G. Kachanoski, *J. Soil Sci.* **63**, 607 (1988).
3. H. Oskarsson, O. Arnalds, J. Gudmundsson, G. Gudbergsson, *Catena* **56**, 225 (2004).
4. R. Lal, *Science* **304**, 1623 (2004).
5. R. Lal, *Environ. Intl.* **29**, 437 (2003).
6. J. C. Ritchie, G. W. McCarthy, *Soil Till. Res.* **69**, 49 (2003).
7. W. H. Renwick, S. V. Smith, R. O. Sleezer, R. W. Buddemier, *Science* **305**, 1567 (2004).
8. R. F. Stallard, *Global Biogeochem. Cycles* **12**, 231 (1998).
9. S. V. Smith, W. H. Renwick, R. W. Buddemier, C. J. Crossland, *Global Biogeochem. Cycles* **15**, 697 (2001).
10. K. Van Oost, G. Govers, T. A. Quine, G. Heckrath, *Science* **305**, 1567 (2004).
11. B. Schlunz, R. R. Schneider, *Int. J. Earth Sci.* **88**, 599 (2000).
12. E. A. Romankevich, A. A. Vetrov, M. E. Vinogradov, V. I. Vedemikov, *Oceanology* **40**, 335 (2000).
13. P. A. Raymond, J. E. Bauer, *Nature* **409**, 497 (2001).
14. P. Jacinthe, R. Lal, *Land Degrad. Dev.* **12**, 329 (2001).
15. D. Pimentel, *Environ. Dev. Sust.* **8**, 119 (2006).
16. N. Borlaug, *Science* **318**, 359 (2007).

Response

WE AGREE THAT ACCELERATED SOIL EROSION should not be promoted as an effective sink for atmospheric CO_2, and we want to emphasize the need for soil erosion control and restoration. However, our results do not support Lal and Pimentel's assertion that agricultural soil erosion is a strong source of atmospheric CO_2.

First, we demonstrated that the carbon content of eroding soils is subject to continual decline as carbon is exported from the site of erosion. It is true that selective erosion of the fine fraction may further accelerate this export. However, severe soil erosion will almost always take place as tillage, rill, and/or gully erosion, which is nonselective (*1*). This local decline does not imply that erosion is a source of atmospheric CO_2 because a fraction of the eroded C was replaced by enhanced stabilization of C in the newly exposed soil, resulting in a net uptake of atmospheric CO_2. Although it is true that erosion may reduce crop yields and hence inputs of carbon at sites of erosion, our mass-balance method accounts for all mechanisms occurring (including a yield decrease) and is not based on any assumptions about productivity or soil organic matter (SOM) replacement rate. Furthermore, the replacement rate of SOM was low on eroded sites, resulting in a small rather than a large sink, but a sink nonetheless.

Second, Lal and Pimentel correctly stated that the fate of SOM during the various phases of the erosion-deposition process should be considered, which is what we did in our analysis. However, the affirmation that detachment and transport of SOM inevitably lead to its accelerated decomposition is incorrect. In contrast, consistent with other studies, our results showed that very little net CO_2 release occurred after transport and burial within our microcatchments (*2*–*4*).

Finally, we agree that CH_4 and N_2O from depositional sites warrant further research, but our report focused on the carbon cycle and therefore did not consider N_2O fluxes. The majority of agricultural soils are mineral soils, which are a sink rather than a source for CH_4 (*5*). Although depositional sites in agricultural fields are wetter, it is unlikely that they will often be flooded and reach a low enough redox potential for methanogenesis (*6*).

KRISTOF VAN OOST,[1] JOHAN SIX,[2] GERARD GOVERS,[3] TIM QUINE,[4] S. DE GRYZE[2]

[1]Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium. [2]University of California, Davis, CA 95616, USA. [3]Katholieke Universiteit Leuven, 3001 Heverlee, Belgium. [4]University of Exeter, EX4 4RJ, UK.

References

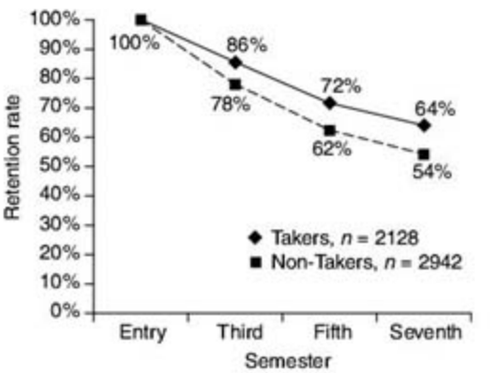
1. P. J. Whiting, E. C. Bonniwell, G. Matisoff, *Geology* **29**, 1131 (2001).
2. K. Van Oost *et al.*, *Global Biogeochem. Cycles* **19**, GB4014 (2005).
3. S. V. Smith, W. H. Renwick, R. W. Buddenmeier, C. J. Crossland, *Global Biogeochem. Cycles* **15**, 697 (2001).
4. G. W. McCarty, J. C. Ritchie, *Environ. Pollut.* **116**, 423 (2002).
5. C. V. Cole *et al.*, *Nutr. Cycl. Agroecosys.* **49**, 221 (1997).
6. D. M. Sylvia, J. J. Fuhrmann, P. G. Hartel, D. A. Zuberer, *Principles and Applications of Soil Microbiology* (Pearson Education Inc., Upper Saddle River, NJ, ed. 2, 2005).

CORRECTIONS AND CLARIFICATIONS

News Focus: "Scientists hope to adjust the president's vision for space" by A. Lawler (1 February, p. 564). Molly Macauley was expressing views offered by participants at last fall's National Research Council forum in Irvine, California, rather than her own opinion.

Random Samples: "When worlds collide" (7 December 2007, p. 1531). The item incorrectly stated that astronomers estimate 1 in 1000 stellar systems host planet-forming dust belts. In fact, that is the number of stars around which astronomers currently see such short-lived belts; many more stars must have them at some time in their history.

Education Forum: "Engineering education research aids instruction" by N. L. Fortenberry *et al.* (31 August 2007, p. 1175). The same data represented by the original figure are regraphed with the *y* axis extended to zero (right).



TECHNICAL COMMENT ABSTRACTS

COMMENT ON "International Conservation Policy Delivers Benefits for Birds in Europe"

Rolando Rodríguez-Muñoz, Alfredo F. Ojanguren, Tom Tregenza

Donald *et al.* (Reports, 10 August 2007, p. 810) assessed the impact of the European Union's Birds Directive, a conservation policy enacted in 1979, and reported evidence for positive population changes in targeted species. We argue that their conclusions are overstatements based on unsuitable data and inappropriate analyses.

Full text at www.sciencemag.org/cgi/content/full/319/5866/1042b

RESPONSE TO COMMENT ON "International Conservation Policy Delivers Benefits for Birds in Europe"

Paul F. Donald, Fiona J. Sanderson, Ian J. Burfield, Stijn M. Bierman, Richard D. Gregory, Zoltan Waliczky

Rodríguez-Muñoz *et al.* raise concerns about our study on the effectiveness of the European Union's Birds Directive, based on the provenance of the trend data used and on the predictions that were tested. Here, we show that our results are robust to different assumptions of uncertainty surrounding trend estimates and that criticisms of the methods stem largely from misunderstandings of the original paper.

Full text at www.sciencemag.org/cgi/content/full/319/5866/1042c

BIOMEDICINE

The Fires of Aging

Donna J. Holmes

The Biology of Human Longevity
Inflammation, Nutrition, and Aging in the Evolution of Lifespans

by Caleb E. Finch

Academic (Elsevier), Amsterdam, 2007. 640 pp. $69.95, £39.99, €58.95. ISBN 9780123736574.

Metabolically speaking, we're all on fire. Current thinking in the biology of aging suggests that the normal processes cells use to burn fuel, providing energy for life, indirectly lead to much of the disease and disability that characterize aging in humans and other animals. Chemically unstable by-products of cellular oxidation—especially free oxygen radicals—can initiate the deterioration of cell membranes and macromolecules (*1*, *2*). As small "hits" causing cellular injury accumulate, the results can range from uncorrected mutations and cancers to forms of tissue damage leading to vascular pathology and Alzheimer's disease. With heart disease, stroke, and diabetes now among the leading causes of aging-related death in the United States (*3*), and rising fast as baby boomers join the ranks of the elderly, these diseases are a primary target for biomedical research.

In *The Biology of Human Longevity*, Caleb Finch spotlights the relationships among cellular processes that are considered important causes of aging-related disease. This formidable book successfully integrates the "free-radical damage" hypothesis with a new more general theory of aging having direct implications for preventing specific pathological syndromes that increase with chronological age. Finch (a professor of gerontology at the University of Southern California) weaves complex strands of evidence from medical history, molecular biology, and biochemistry (as well as evolution and anthropology) into a layered tapestry of the synergies among various forms of "bystander damage" (injury, mutations, toxic insult, and oxidative damage) and the ways in which these can foreshadow diagnosable disease states. Oxidative damage remains a central player in the drama he unfolds, but now it shares the stage with several lesser-known, equally important accomplices: inflammation, damage during development, and the hazards of overnutrition.

Finch's central thesis is wedded to a fundamental tenet of evolutionary biology: physiological adaptations for sustained health and reproduction involve evolutionary trade-offs. The growth and maintenance of reproductive and somatic tissues, for example, require hormones and growth factors (e.g., sex steroids and insulin-like growth factors) that also facilitate the proliferation of cancers and other pathological changes. Effective defenses against infection, injury, or stress—even during prenatal development—ignite a cascade of inflammatory factors, including cytokines and other peptides. This first wave of the innate immune response is followed by rapid proliferation of specialized cells and the release of a plethora of signaling molecules and oxidative by-products that can damage cells and macromolecules.

Adding fuel to the fire is a diet too rich in animal fat, increasing exposure to pathogenic microbes and exacerbating inflammation. The book's cover features a photo of the "Venus of Willendorf," an exceedingly curvaceous, female stone effigy (~23,000 BCE). In a world of nutritional scarcity, she represented an unrealistic paragon of reproductive potential. Today, we consider a 1.65-m-tall woman of similar proportions—say, a body mass of 82 kg—to be obese (*4*), with clinically elevated risks for type 2 diabetes and cardiovascular disease.

The "Venus of Willendorf" effigy (Naturhistorisches Museum, Vienna).

CREDIT: ARCHIVO ICONOGRAFICO, S.A./CORBIS

In his last and most provocative section, Finch proposes that increases in brain size and the human life span over the past million years occurred in concert with changing nutritional priorities, slower developmental rates, and a tolerance for inflammation in "dirty, invasive, and stingy" prehistoric environments. The integration of more meat into the human diet, he argues, provided protein needed for larger brains but involved new physiological and genetic trade-offs between fitness and liability for long-term damage. This scenario provides a satisfying rationale for why variants of some genes for metabolizing animal fat that are linked to a human predilection for atherosclerosis, some cancers, and the amyloid plaques characteristic of Alzheimer's disease (such as those of the ApoE gene family) are not shared by our closest primate relatives.

The book provides an unparalleled synthesis of the burgeoning literature addressing the roles played by oxidative damage and inflammation in diseases of aging. Finch offers thorough discussions of methodologies and central constructs in aging research (including the search for better molecular biomarkers of aging) and a short primer of evolutionary aging theory—possibly not enough for some evolutionary biologists, but a sufficiently broad perspective for those wading into the field for the first time. Along with a catalog of experimental and genetic interventions in aging based on laboratory animal systems, he summarizes an exhaustive body of clinical literature, showcasing nutritional and pharmaceutical approaches that have potential for intervening in aging disease states by curtailing inflammation-induced damage.

With the coupling of his expertise in neuroscience and clinical medicine to his keen interests in demography and comparative zoology, Finch arguably remains our most potent synthesizer of biology and gerontology. Here his writing conveys a sense of urgency not present in his classic *Longevity, Senescence, and the Genome* (*5*). A bit verbose and convoluted at times, *The Biology of Human Longevity* would have benefited from the attention of a copy editor half as ambitious as the author. That aside, the intellectual framework Finch provides in it will be intensely stimulating to both experts and newcomers in the field of aging.

The reviewer is in the School of Biological Sciences and Center for Reproductive Biology, Washington State University, Pullman, WA 99164–4236, USA. E-mail: djholmes@wsu.edu

The book should attract notice from developmental biologists, anthropologists, and clinical researchers alike.

References and Notes

1. D. Harman. *J. Gerontol.* **11,** 298 (1956).
2. K.B. Beckman, B. N. Ames. *Physiol. Rev.* **78**, 547 (1998).
3. R. N. Anderson, B. L. Smith, *Natl. Vital Stat. Rep.* **53** (no. 17) (2005); www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_17.pdf.
4. See the NIH's calculator for body mass index at www.nhlbisupport.com/bmi.
5. C. E. Finch, *Longevity, Senescence, and the Genome* (Univ. of Chicago Press, Chicago, 1990).

10.1126/science.1153829

PHILOSOPHY

Finding Freedom Through Complexity

Todd Stuart Ganson

Since the 17th century one of the central problems of philosophy has been to make sense of certain phenomena that lack any obvious place in our modern scientific view of the world—phenomena like colors as we experience them, norms of thought and action, and free and responsible action. Nancey Murphy and Warren S. Brown's fascinating *Did My Neurons Make Me Do It?* touches on all of these phenomena, but its focus is on the third, on the difficult problem of understanding how such action might be possible in light of our growing scientific knowledge about humans and the world they live in.

Did My Neurons Make Me Do It?
Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will

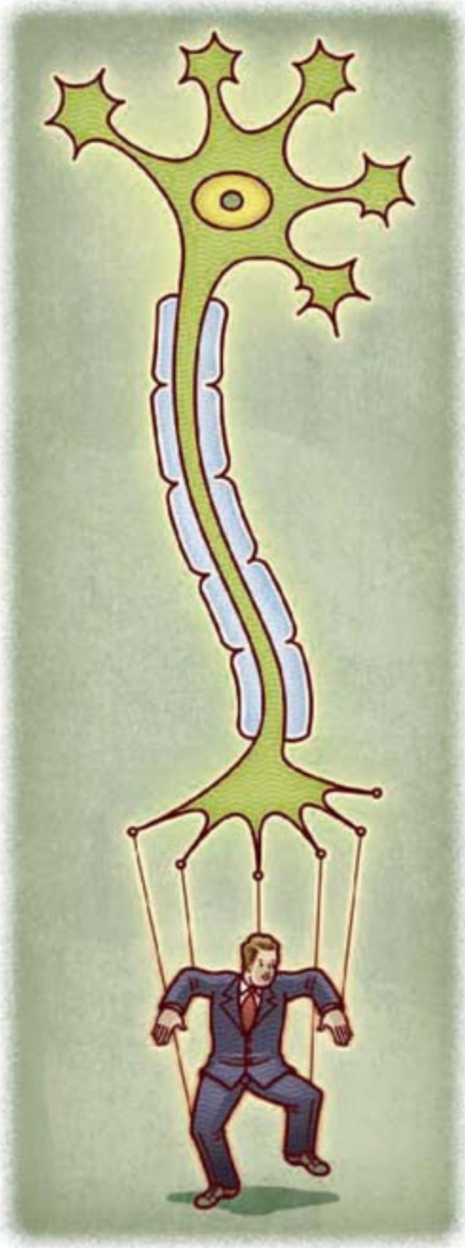
by Nancey Murphy and Warren S. Brown

Oxford University Press, Oxford, 2007. 352 pp. $65, £40. ISBN 9780199215393.

When accounting for mental phenomena, philosophers typically wish to avoid the extremes of either outright eliminativism (the claim that many or all of the mental states posited by common sense are nonexistent) or supernaturalism. One middle ground popular among philosophers is to endorse a reductive view of the mental, the view that our mental lives are fully explicable in terms of our neurochemistry. This sort of view will often go hand-in-hand with the claim that mental states are nothing other than certain chemical states of the brain—the mind just is the brain. Although this way of thinking about the mind fits nicely with the perspective of modern science, it appears to have the consequence that our choices and actions are fixed or determined by the laws of chemistry. And in that case there seems to be an important sense in which we are unable to do otherwise than what we in fact do. Because acting freely (in contrast with acting voluntarily) presupposes the ability to do otherwise, this reductive view of the mental evidently avoids eliminativism about the mental at the cost of denying the existence of genuine freedom.

There are several familiar ways to try to avoid this consequence that we lack freedom. One is to suggest that perhaps the laws governing the activity of our neurons are statistical in character and allow for genuine randomness. This suggestion by itself is inadequate because it is perhaps equally difficult to reconcile freedom with randomness or chance. A more common response is to insist that having freedom and responsibility is perfectly compatible with being determined at the neurological level. Murphy and Brown (professors of, respectively, philosophy and psychology at Fuller Theological Seminary, Pasadena, California) do not engage much with this traditional debate between compatibilists and incompatibilists. Instead, they focus on developing an alternative to the reductive view of the mental, an alternative that allows for top-down or emergent causation. Get rid of the reductive view of the mental, and you get rid of the main source of the worry that our actions are determined by laws of chemistry.

It is difficult initially to make sense of emergent causation. Suppose that a mental state or event M is realized by some neural event N and we want to say that M causes, say, efferent nerve stimulation N*. Presumably M causes N* only because there is a lawful connection between N and N*. But in that case it is difficult to see what causal work M is doing. We want a plausible alternative to the reductive approach that avoids falling into some form of epiphenomenalism. Murphy and Brown suggest that complex systems can causally influence what bottom-level events occur by shaping the conditions that trigger those events. This suggestion has the advantage that it makes emergent causation seem less mysterious. Statements of causal laws will typically include some reference to the conditions under which the law is operative, and it would seem that structural features of a system might themselves serve as conditions that influence component, lower-level processes.



CREDIT: PETER HOEY

Murphy and Brown take their move away from the reductive approach to involve something akin to a paradigm shift. They do not engage in a detailed criticism of reductionist views. Rather, they sketch in broad outline ways of thinking about cognition, motivation, and language that are more in line with an emergentist point of view. Their discussion is rich in references to the relevant empirical research and a valuable resource for those interested in the literature on emergentism. The discussion is a bit one-sided at times, and it would have benefited from a more rigorous and charitable engagement with reductionist views. Nevertheless, *Did My Neurons Make Me Do It?* is a nicely written, engaging book that makes a genuine contribution to the growing literature on mental causation.

10.1126/science.1153669

The reviewer is at the Department of Philosophy, King 105, Oberlin College, 10 North Professor, Oberlin, OH 44074–1019, USA. E-mail: todd.ganson@oberlin.edu

ECONOMICS

The Demography of Educational Attainment and Economic Growth

Complementing primary education with secondary education in broad segments of the population is likely to give a strong boost to economic growth.

Wolfgang Lutz,[1]* Jesus Crespo Cuaresma,[2] Warren Sanderson[3]

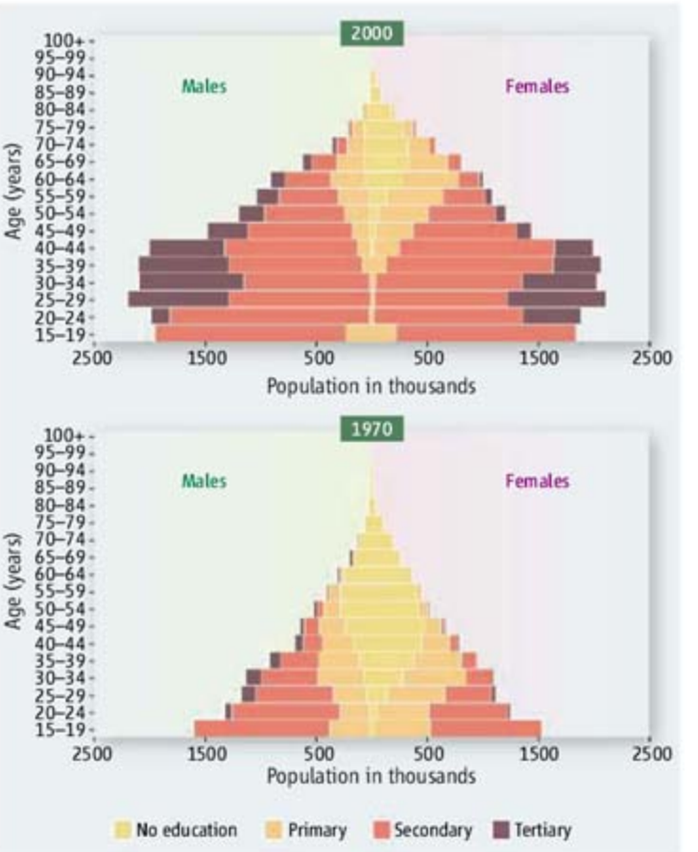
When the world leaders convened in New York in 2000 and solemnly announced the United Nations (U.N.) Millennium Development Goals (MDGs), the goal of universal primary education by 2015 figured prominently, second only to the reduction of extreme poverty and hunger. Although the diminution of poverty and hunger is a self-evident end in itself, putting all children into school is a goal primarily because it is believed to be the vehicle through which the level of adult human capital is improved, in turn enhancing individual well-being, health, and economic growth. The empirical basis for assuming an important positive effect of education on economic growth is, however, surprisingly weak. Although it is well established that, at the individual level, more years of schooling lead to higher income, at the macroeconomic level, the empirical evidence, so far, relating changes in education measures to economic growth has been ambiguous.

The MDG's focus on universal primary education has to be seen in the context of the evolution of international concerns about eradicating illiteracy with a series of well-intended but "demographically illiterate" and therefore unrealistic goals, which ultimately failed. In 1990 at a historical conference in Jomtien (Thailand), 155 governments and 150 organizations issued a World Declaration on Education for All that included the prominent goal of reducing the adult illiteracy rate to one-half of its 1990 level by the year 2000. From a demographic perspective, it is clear that this goal is impossible to achieve in poorly educated countries if education is primarily concentrated in young people, because it would take many decades for the better-educated youngsters to replace the illiterate adult population. Ten years later in Dakar, the international community, confronted with the failure of the previous goals, chose to use an even more unfortunate formulation of their goal, namely, "increase the literacy rate by 50%," which, for countries that already had more than 66.7% literacy, implied a goal of more than 100%. To correct this evident political innumeracy, the U.N. Educational, Scientific, and Cultural Organization (UNESCO) later modified the goal to imply a goal of 100% for countries that already had 66.7% literacy or more. Leaving aside the fact that 100% literacy may be impossible to reach for any society, even this revised goal focuses on the stock of adult literacy without considering the cohorts involved.

The MDG's focus on enrollment in formal primary education is related but somewhat different from the earlier focus on literacy: Literacy is a skill that, in principle, can also be acquired outside the regular education system and that can be lost again (secondary illiteracy). In contrast, once a certain level of educational attainment is reached, it cannot be lost again throughout the rest of our lives. We will be able to keep our doctorates even if our skills seriously degrade; this property actually makes educational attainment distributions easier to model because movements can only go in one direction, toward higher education.

Using the demographic method of multistate back projection, a group of researchers at the International Institute for Applied Systems Analysis (IIASA) and the Vienna Institute of Demography (VID) has recently completed a full reconstruction of educational attainment distributions by age and sex for 120 countries for the years 1970–2000 (*1*). This improvement in human capital is illustrated (see figure, below) along cohort lines in the case of South Korea. The advantages of this data set relative to others (*2*–*4*) are its detail (four educational categories for 5-year age groups of men and women), its consideration of differential mortality, and its strict consistency of the definition of educational categories over time. The age and education composition detail in our new data allow us to perform more detailed statistical analyses of the relation between education and economic growth than can be performed using preexisting data.

Previous cross-country economic growth regressions tended to show that changes in educational attainment are largely unrelated to economic growth [for example, (*5*, *6*)], which contradicts theory and microeconometric evidence. Most of the literature in this field attributes the existence of this puzzle to



Reconstructing educational attainment. (**Top**) Age pyramid of South Korea in 2000 with colors indicating different educational attainment categories. (**Bottom**) Reconstructed age pyramid of South Korea in 1970 with colors indicating different educational attainment categories.

[1]Vienna Institute of Demography (Austrian Academy of Sciences), A-1040 Vienna, and World Population Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria. [2]University of Innsbruck, A-6020 Innsbruck, and World Population Program, IIASA. [3]Stony Brook University, Stony Brook, NY 11794, USA, and World Population Program, IIASA.

*Author for correspondence. E-mail: lutz@iiasa.ac.at

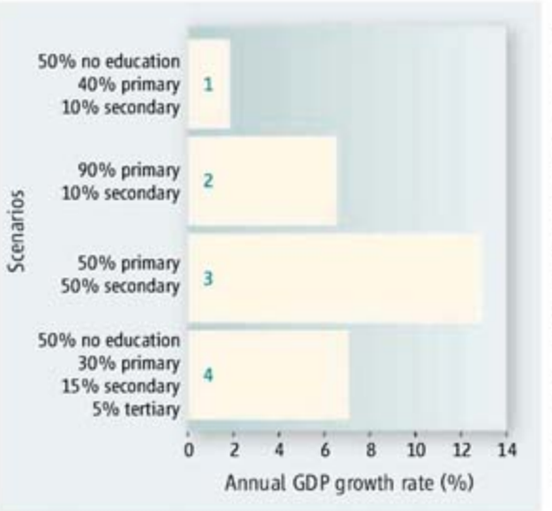
deficiencies in the series of education data (*3*, *4*, *7*). Also, averaging education attainment data over longer horizons has led to more consistent patterns (*8*), which highlights the importance of measurement problems in the previous data.

Using our new educational attainment data by age groups, we estimated simple growth regressions based on 5-year periods for a panel of 101 countries for which all the necessary economic and education data exist over the period 1970–2000. It has a fair representation of all continents (*9*). These new data allow us to use the education levels of different age groups as potential determinants of economic growth. The results show consistently positive, statistically significant education effects on economic growth for some age and education groups (*9*) and, hence, make the puzzle disappear.

In our model, human capital by broad age groups enters production both as differentiated labor force inputs and through the absorption rate of new technologies, which, in turn, depends on the interaction between human capital and distance to the technological frontier [see (*5*) for a similar approach]. The effect of education on labor force participation is assumed to be specific to each age and education group and constant over time. The model is described, related to the literature, and justified in detail (*9*).

The implications of the results and the value added by the new data set are illustrated by simulating four scenarios based on estimated coefficients (see figure below, table S1, and the discussion in the supporting online text). These four roughly resemble alternative hypothetical education policy strategies for a poor African country. In these simulations, we focus on the fact that economic growth is determined by the adoption of new technologies, and that the size of the effect depends on the income level of the country under study, because countries that are further away from the technology frontier are able to profit more [in terms of Gross Domestic Product (GDP) growth] from this channel of the education-growth link.

The figure above presents the average annual GDP growth rates corresponding to different education level distributions. Scenario 1 presents the reference case of a country with a young age structure (70% of the population in the 15- to 40-year-old group and 30% in the 40- to 65-year-old group), a low starting level of income and investment rate and the following educational structure: half of the population without any formal schooling, 40% with some primary and 10% with at least completed junior secondary school (but no tertiary education). The education groups used in the analysis (no education, primary, secondary, and tertiary) are nonoverlapping. This roughly fits the demographic structure of some Latin American and African countries in our sample, e.g., Guatemala, Honduras, Kenya, Rwanda or Uganda. On the basis of the estimated model, such a country would have rather slow economic growth. Scenario 2 considers the otherwise identical country under the hypothetical assumptions that it has for long met MDG goal 2 and that the previously uneducated half of the adult population now has primary education. This case would lead to somewhat higher average growth of GDP. Scenario 3 considers a possible new MDG effort that adds widespread secondary education (we assume here 50% of the population achieving at least some secondary schooling) to universal primary. The model simulations indicate that this additional investment in secondary education provides a huge boost to economic growth, over five times the level of the baseline scenario and also much more than in the scenario of universal primary education alone.



Annual GDP growth rates according to the four alternative educational attainment distributions (see text).

Scenario 4 finally presents another possible direction of improvement from the baseline (which somewhat resembles the case of India), in which half of the population remains without education although 5% have tertiary education, 15% secondary, and 30% primary. This case of elitist education in a context with half of the population being without any schooling does clearly better than the baseline and even better than the universal primary education (combined with 10% secondary and no tertiary), but falls far short of the economic growth implied by universal primary combined with 50% secondary and no tertiary education.

We compared these results with an age-aggregated version of the IIASA-VID data and the widely used Barro-Lee (*2*) data set, which has no age detail (see table S2). The comparison with the full age-structured model gives evidence of differences across age groups in the effects of education on GDP growth (table S1). These results point to the importance of the demographic structure of human capital when assessing the effect of education on economic growth. The IIASA-VID data set is, as of today, the only comprehensive data set offering such demographic detail in education figures.

These new findings have political consequences for the next round of defining international education goals (*10*): The current MDG's focus on universal primary education is important but insufficient. It needs to be complemented with the goal of giving broad segments of the population at least a completed junior secondary education. Only this is likely to give initially poor countries the human capital boost that is necessary to bring large segments of the population out of poverty. For more industrialized countries, tertiary education of younger adults also is an important determinant of economic growth.

In conclusion, better education does not only lead to higher individual income but also is a necessary (although not always sufficient) precondition for long-term economic growth. The fruits of investments in education need a long time to ripen, to translate the education of children into better human capital of the adult labor force. Education is a long-term investment associated with near-term costs, but, in the long run, it is one of the best investments societies can make in their futures.

References and Notes

1. W. Lutz, A. Goujon, Samir K.C., W. Sanderson, *Vienna Yearb. Popul. Res. 2007*, 193 (2007); http://hw.oeaw.ac.at/1728-4414?frames=yes.
2. R. J. Barro, J. W. Lee, *Am. Econ. Rev.* **86**, 218 (1996).
3. A. de la Fuente, R. Domenech, *J. Eur. Econ. Assoc.* **4**, 1 (2006).
4. D. Cohen, M. Soto, *J. Econ. Growth* **12**, 51 (2007).
5. J. Benhabib, M. Spiegel, *J. Monet. Econ.* **34**, 143 (1994).
6. L. Pritchett, *World Bank Econ. Rev.* **15**, 367 (2001).
7. A. Krueger, M. Lindahl, *J. Econ. Lit.* **39**, 1101 (2001).
8. R. Topel, *Handb. Labor Econ.* **3**, 2943 (1999).
9. Additional information is provided in the supporting material available on *Science* Online.
10. J. Cohen, D. E. Bloom, M. B. Malin, Eds., *Educating All Children: A Global Agenda* (American Academy of Arts and Sciences, Cambridge, MA, 2007).
11. The input of A. Fürnkranz-Prskawetz, A. Goujon, Samir K.C., and S. Scherbov is gratefully acknowledged.

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5866/1047/DC1

10.1126/science.1151753

CANCER

A Skin Cancer Virus?

High-throughput DNA sequencing has identified a new human virus in a rare but aggressive form of skin cancer.

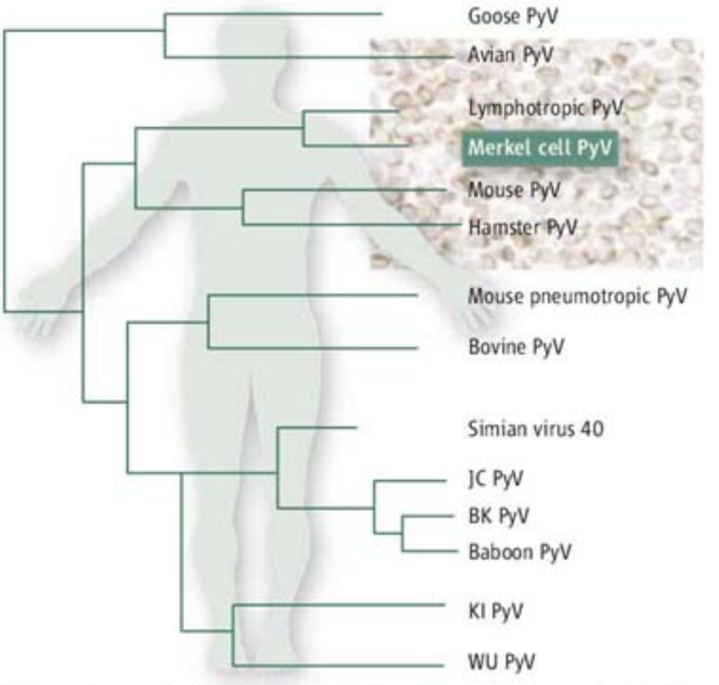
Raphael P. Viscidi[1] and Keerti V. Shah[2]

Is there a virus that causes Merkel cell carcinoma, a rare but aggressive skin cancer? Maybe. On page 1096 in this issue, Feng *et al.* (*1*) describe Merkel cell polyomavirus, a new human virus that is associated with a particular neuroectodermal skin cancer that occurs primarily in elderly immunosuppressed individuals. It may be the first example of a human cancer caused by a polyomavirus, a viral family whose association to other cancers has been controversial.

Human polyomaviruses BKPyV and JCPyV were discovered in 1971, and are common infections of the urinary tract, responsible, respectively, for nephropathy in kidney-transplant recipients and progressive multifocal leukoencephalopathy in immunosuppressed populations. Both viruses are widespread and persist as latent infections without causing disease. However, in immunocompromised patients, they can become pathogenic. In 2007, the genomes of two additional human polyomaviruses, KIPyV and WUPyV, were found by molecular screening of patients' respiratory tract samples (*2*, *3*), but we know little about their epidemiology and pathogenic potential.

Merkel cell carcinoma metastasizes quickly, and like other types of skin cancer, its incidence is increasing worldwide. To investigate whether this cancer may have an infectious origin, Feng *et al.* constructed two expression libraries, one from a single tumor and the other from a pool of three tumors, and examined them by digital transcriptome subtraction. This method quickly reveals sequence differences, and in this case, was used to detect nonhuman sequences. In the library from the single tumor, they found a transcript corresponding to a fusion between a polyomavirus T antigen–like sequence and part of the coding sequence of a human receptor tyrosine phosphatase gene on chromosome 3p14.2. Using a variety of techniques and additional tumor tissues, the authors then identified the complete sequence of the viral genome and showed that it was integrated into tumor cell DNA in six of eight virus-positive tumors. The patterns of integration observed at different chromosomal sites suggest that viral infection precedes clonal expansion of the tumor cells. Of the four tumors sampled by digital transcriptome subtraction, one had a high copy number of transcripts containing viral sequence.

To determine whether the polyomavirus has a causal role in the development of Merkel cell carcinoma, Feng *et al.* analyzed tissue samples from 10 patients with Merkel cell carcinoma, and from 84 patients without the cancer. Genomic sequences containing Merkel cell polyomavirus were unequivocally detected in tumors from 8 of 10 patients with the carcinoma, as well as in a lymph node to which the cancer had metastasized, from a patient with a virus-positive tumor. None of the 84 samples from patients without Merkel cell carcinoma were unequivocally positive.



Polyomaviruses. A schematic representation of the evolutionary relationships among the animal and mammalian polyomaviruses. KI, WU, JC, BK, and Merkel cell polyomaviruses infect humans. (**Inset**) Merkel cell carcinoma, cytokeratin 20 staining; magnification ×40 (*1*).

Phylogenetic analysis reveals deep evolutionary relationships among the mammalian polyomaviruses (see the figure) (*4*), and supports the existence of four clades represented by mouse, bovine, and simian/ human polyomaviruses (which includes BKPyV and JCPyV), and a separate clade formed by the newly discovered KIPyV and WUPyV. The latter three are sister clades that share a common ancestor, and are distantly related to the mouse polyomavirus clade. Phylogenetically, Merkel cell polyomavirus falls within the mouse polyoma clade and is most closely related to the lymphotropic polyomavirus (also known as African green monkey polyomavirus) of presumed simian origin. Polyomaviruses are highly species-specific and are believed to evolve in close association with their host. The evolutionary position of Merkel cell polyomavirus calls into question this concept and may indicate that host switching can occur.

Polyomaviruses are oncogenic when inoculated into experimental animals, and the viral proteins expressed early in the infectious cycle initiate transformation and immortalization of cultured cells. Therefore, the possible role of BK and JC polyomaviruses in the etiology of human tumors has been extensively investigated, but the results have been inconclusive. Is Merkel cell polyomavirus the "cause" of Merkel cell carcinoma? Merkel cell carcinoma is the first example of a human cancer in which a polyomavirus genome is naturally integrated into cellular DNA, the key event in experimental polyomavirus-induced animal tumors.

Nevertheless, a number of questions must be answered before making any judgment regarding causation. The two complete Merkel cell polyomavirus genomes sequenced by Feng *et al.* have truncating mutations in the carboxyl terminus of the large T antigen open reading frame. It is not clear whether the encoded viral large T protein possesses the transforming and immortalizing properties of other polyomavirus large T antigens. We also don't know whether viral oncoproteins are expressed in all virus-positive tumors. The

[1]Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. E-mail: rviscid1@jhmi.edu, [2]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. E-mail: kvshah@jhsph.edu

finding of viral genome integration into the host genome in this cutaneous cancer must also be reproducible in other studies. BK and JC polyomaviruses are common infections of childhood and are pathogenic only when they are reactivated in immunosuppressed individuals. Are Merkel cell polyomavirus infections similarly related to the pathogenesis of a subset of Merkel cell carcinoma? We need to learn about the natural history of Merkel cell polyomavirus infection in humans before we can accurately assess its etiologic link to Merkel cell carcinoma in the framework proposed by Hill (*5*)—by examining virus-tumor association for its strength, consistency, specificity, plausibility, and temporality.

References

1. H. Feng, M. Shuda, Y. Chang, P. S. Moore, *Science* **319**, 1096 (2008); published online 17 January 2008 (10.1126/ science.1152586).
2. T. Allander *et al.*, *J. Virol.* **81**, 4130 (2007).
3. A. M. Gaynor *et al.*, *PLoS Pathog* **3**, e64 (2007).
4. M. Pérez-Losada *et al.*, *J. Virol.* **80**, 5663 (2006).
5. A. B. Hill, *Proc. R. Soc. Med.* **58**, 295 (1965).

10.1126/science.1155048

MATERIALS SCIENCE

The Cutting Edge of Plasma Etching

Thorsten Lill[1] and Olivier Joubert[2]

Better etching methods for creating more finely structured microcircuits will require close attention to basic plasma physics and surface chemistry.

Plasma etching is the universal tool to define high-resolution patterns in integrated chip manufacturing. To create structures in a chip, a pattern is formed in a photoresist by lithography and then transferred into the device materials by plasma etching. As the industry approaches the ability to create microcircuit structures on the order of 20 nm, this technology faces fresh challenges. To make progress, we need to go back to the basic science of how plasmas interact with surfaces.

Several trends are at work: First, circuit patterns need to be accurate to within 1 nm and below, within a single wafer and across several wafers. Second, plasma etching is becoming an integral part of pattern generation (using lateral erosion of the lithographic photoresist to improve resolution, for example). Third, aspect ratios of the final structures (that is, the ratio of length to width) are increasing dramatically. Finally, the number of potential new material candidates and their possible combinations in future structures is exploding.

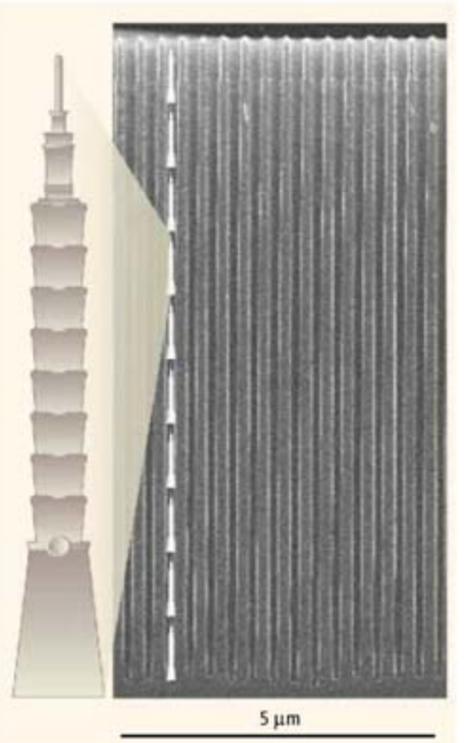
To meet these challenges, the range of plasma densities and chemistries and the ion energies that can be generated must be continuously increased. Whereas the last decade was mainly characterized by the search for the best plasma sources for the different etch applications, the focus today is on making smaller structures with extreme uniformity across the wafer. State-of-the-art etch reactors must allow for perfect center-to-edge uniformity of neutral atoms and ions in the plasma, and uniform wafer surface temperature for any plasma regime that can be generated.

The other frontier of advanced etch reactor development is the extension of the range of usable ion energies. At one extreme, ion energies are being extended to very low values—for instance, tens of electron volts for high-κ (dielectric constant) materials etching. On the other end of the ion energy spectrum, high–aspect ratio oxide etching requires extremely high ion energies (thousands of electron volts). The introduction of novel materials such as high-κ dielectrics is also driving the development of plasma etching at elevated wafer temperatures.

Historically, there was a sharp dividing line between reactors used for oxide etching on the one hand and reactors for silicon and metal (conductor) etching on the other. Process engineers have preferred capacitively coupled plasmas (CCPs) for oxide etch, but inductively coupled plasmas (ICPs) or electron cyclotron resonance (ECR) plasmas for conductor etch. However, the lines are now blurred and both types (CCP and ICP) can be used for a variety of applications.

To create the various necessary microstructures, a new line of division is being drawn: high-fidelity patterning for moderate aspect ratios (up to 10:1), and etching with very high aspect ratios (up to 100:1; see the figure). The latter requires very high ion energies and is the domain of the CCP machines, whereas the former is more suited for ICP reactors. As a result, we have to better understand the surface reaction mechanisms occurring on the wafer and chamber walls during plasma etch to select the right reactor for each application.

As integrated circuit manufacturers try to differentiate their products, they compete by using new materials and device designs. Plasma etch is very much at the center of these efforts. This trend requires very fast adaptation of well-characterized fabrication processes. In this context, understanding the fundamentals of the etch mechanisms and their correlations to key process parameters is crucial. Each etch step must be characterized not only by etch rate and uniformity, but also by more fundamental properties such as the composition, thickness, and line-edge roughness of the sidewall layers of the structure; the chemical nature of etch by-products deposited on the chamber walls (which affects process stability and reproducibility); the thickness of the etch-front mixing layer (correlated to etch rate and selectivities between layers); and the impact of aspect ratio–dependent etching phenomena. This systematic approach is extremely critical as these reactors increase in complexity and must handle an ever-increas-



In the trenches. The trench-like microstructures in advanced dynamic random access memories reach aspect ratios of 90 or more. This image shows trenches with aspect ratios of 80:1. The Taipei 101 tower, scaled down to the trench width, would fit 10 times into this trench. On one wafer, there are about 7×10^{11} trenches with a combined length of 4000 km. Achieving the high precision and uniformity required of such structures will demand better understanding of how plasmas interact with surfaces.

CREDIT: ADAPTED FROM THORSTEN LILL AND OLIVIER JOUBERT

[1]Applied Materials Inc., Santa Clara, CA 95054, USA. [2]Laboratoire des Technologies de la Micro-électronique–Centre Nationale de la Recherche Scientifique, 38054 Grenoble, France. E-mail: thorsten_lill@amat.com

ing number of very specific etch applications.

Obtaining identical process performance from one chamber to another ("chamber matching") and process repeatability are quickly becoming the biggest challenge for plasma etching. Because complex microstructures contain a larger number of thinner layers, the etch recipes contain multiple short etch steps. The result is an increase in the amount of time during which the reactor is in a state of transition from one processing condition to the next. Advanced etch recipes to pattern a transistor electrode contain 5 to 10 steps with a total of 100 to 150 parameters, which demands repeatable step-to-step transition.

In addition, chamber walls have an effect on complex recipes with multiple steps and chemistries. A deep knowledge of the coatings formed on the chamber walls during the process and the capability to modify them is needed to master this challenge (*1*). Because of all the variables and tuning that can be done, nearly perfect calibration is required. When companies launch the final etch process into mass manufacturing, process or hardware changes to address matching issues can be extremely costly. Achieving chamber matching will enable the right choices to be made at the beginning.

The evolution of plasma etching has reached a point where the fine-tuning in hardware design of the etch reactor and process parameters must be always accompanied by a strong understanding of the fundamentals. Plasma etching should remain the universal patterning tool in the foreseeable future as long as chamber matching can be ensured. Plasma etching in highly complex reactors might become supplemented with some alternative means of patterning—neutral beam etching, atomic layer etching, and other methods are being discussed—but it is not likely that these technologies will become as universal in their capabilities. The synergy of chemical and physical etching is still the most versatile tool for patterning of integrated chips (*2*).

References and Notes

1. R. Ramos *et al.*, *Thin Solid Films* **515**, 4846 (2007).
2. J. Coburn, H. Winters, *J. Appl. Phys.* **50**, 3189 (1979).
3. We thank M. Oswald for the calculation of trench length on a wafer.

10.1126/science.1153901

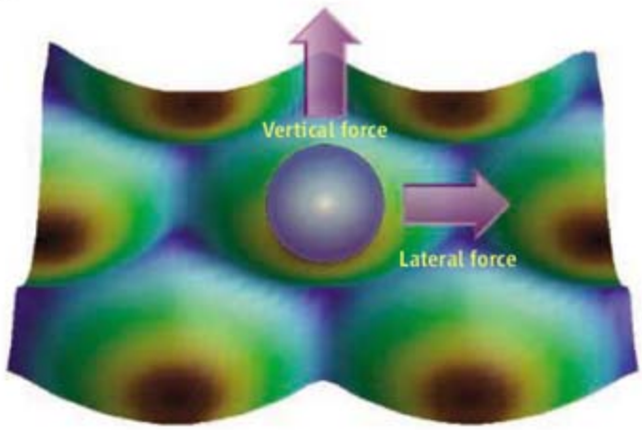
MATERIALS SCIENCE

How to Move an Atom

An instrument that combines the strengths of AFM and STM allows determination of the forces required to move a single atom on a surface.

Oscar Custance[1] and Seizo Morita[2]

Knowledge of the forces required to lift up, pull, or push objects and building blocks is of fundamental importance in many fields from engineering to biotechnology (*1*). On page 1066 of this issue, Ternes *et al.* report measurements of the vertical and lateral forces to pull one of the smallest building blocks: a single atom (*2*). They accomplished this task by combining the two most powerful techniques for interacting with matter at the nanometer scale—scanning tunneling microscopy (STM) and atomic force microscopy (AFM)—in a single instrument.



Atomic manipulation. Using a sophisticated scanning probe implementation, Ternes *et al.* determine how much lateral and vertical force is required to pull an atom on a metallic surface and how these forces are related to each other.

In 1990, Eigler and Schweizer first assembled a nanometer-scale structure, atom by atom, on a surface (*3*). Following this feat, scientists began to envision nanometer-scale devices based on a few atoms and materials engineered at the atomic scale. Important steps toward these concepts have been accomplished with STM; for example, scientists have studied the initial states of "wiring" a single molecule (*4*), built a logic gate based on molecules (*5*), and controlled chemical reactions at the atomic level (*6*). More recently, AFM has been used to manipulate atoms at cryogenic temperatures (*7*) and to create complex nanostructures, one atom at a time, at room temperature (*8*).

The manipulation mechanisms have been investigated with both techniques (*9*, *10*), but the lateral forces applied in atomic manipulation experiments have not yet been measured accurately. To address this issue, Ternes *et al.* updated the STM built by D. M. Eigler with a qPlus sensor (*11*). This scanning probe implementation is based on a small quartz tuning fork, with one of the prongs fixed and the other bearing a sharp tip at its free end. The free prong acts as a cantilever; variations in its longitudinal bending can be detected by measuring tiny piezoelectric currents produced when the quartz deforms.

The authors operate the qPlus sensor using the frequency modulation detection method (*12*). They oscillate the free prong at resonance, keeping the oscillation amplitude constant. The presence of a force between tip and surface produces a variation (Δf) in the resonant frequency of the prong. For sufficiently small oscillation amplitudes, Δf is a rough approximation of the tip-surface interaction force gradient [k_z in (*2*)]. By recording Δf as a function of the tip-sample separation, subtle tip-surface interaction forces can be quantified (*13*, *14*) by inverting the complex mathematical expression that predicts the behavior of Δf (*15*). An interesting property of the qPlus sensor is its elevated static stiffness (k_s, the force required to bend the free prong), with typical values 40 times that of the stiffest silicon cantilevers commercially available for AFM. These high k_s values enable the authors to reduce the oscillation amplitude to a few tens of picometers; under these conditions, the tunneling current flowing between tip and surface can be reliably measured simultaneously with Δf.

The authors measure these magnitudes in consecutive constant-height line scans that get progressively closer to the surface until they reach the threshold for atomic manipulation, and quantify the tip-surface interaction force in

[1]National Institute for Materials Science (NIMS), 1-2-1 Sengen, 305-0047 Tsukuba, Ibaraki, Japan. E-mail: custance.oscar@nims.go.jp [2]Osaka University, 2-1 Yamada-Oka, 565-0871 Suita, Osaka, Japan. E-mail: smorita@eei.eng.osaka-u.ac.jp

the vertical direction along the line scan. Integration of these forces over different tip-height intervals provides the associated interaction potential, and differentiation of this scalar quantity with respect to the lateral tip displacement allows estimation of the tip-surface lateral forces. In this way, the authors can separately assess the vertical and lateral threshold forces associated with the atomic manipulation.

An unexpected result is that the lateral force to pull a Co atom on a Pt(111) surface seems to be independent of the vertical force (at least in the tip-surface distance range explored by the authors), with a value that remains almost constant. This behavior enables Ternes *et al.* to compare the lateral forces required to move the same atomic species on other metallic substrates; they can thus quantify friction at a single-atom scale.

Applying this approach to full images instead of line profiles, the authors generated maps of the tip-surface interaction potential. These potential maps provide information about the adsorption stability of the atom (or molecule) on the surface.

The work of Ternes *et al.* opens up new avenues for investigations on nanostructuring and studies of friction at the single-atom scale. The possibility of combining the powerful tools provided by STM and AFM in a single instrument, and of correlating tip-surface short-range chemical forces with simultaneous tunneling currents at the atomic and molecular scale, opens a new dimension in nanoscience that will bring exciting discoveries and technological achievements.

References

1. E. A. Galburt *et al.*, *Nature* **446**, 820 (2007).
2. M. Ternes, C. P. Lutz, C. F. Hirjibehedin, F. J. Giessibl, A. J. Heinrich, *Science* **319**, 1066 (2008).
3. D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524 (1990).
4. G. V. Nazin, X. H. Qiu, W. Ho, *Science* **302**, 77 (2003); published online 4 September 2003 (10.1126/science.1088971).
5. A. J. Heinrich, C. P. Lutz, J. A. Gupta, D. M. Eigler, *Science* **298**, 1381 (2002); published online 24 October 2002 (10.1126/science.1076768).
6. S.-W. Hla, L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **85**, 2777 (2000).
7. N. Oyabu, Y. Sugimoto, M. Abe, O. Custance, S. Morita, *Nanotechnology* **16**, S112 (2005).
8. Y. Sugimoto *et al.*, *Nature Mater.* **4**, 156 (2005).
9. L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **79**, 697 (1997).
10. Y. Sugimoto *et al.*, *Phys. Rev. Lett.* **98**, 106104 (2007).
11. F. J. Giessibl, *Appl. Phys. Lett.* **76**, 1470 (2000).
12. T. R. Albrecht, P. Grütter, D. Horne, D. Rugar, *J. Appl. Phys.* **69**, 668 (1991).
13. M. A. Lantz *et al.*, *Science* **291**, 2580 (2001).
14. Y. Sugimoto *et al.*, *Nature* **446**, 64 (2007).
15. J. E. Sader, S. P. Jarvis, *Appl. Phys. Lett.* **84**, 1801 (2004).

10.1126/science.1154853

ECONOMICS

Learning with Regret

Michael D. Cohen

Choices in economic games are predicted better by models that look back at what might have been, instead of looking forward to maximum gain.

In the 1970s and 1980s, researchers in economic psychology such as Herbert Simon challenged the strong assumption made by economic theorists that individual decision-making is purely rational. Economists responded with vigorous new lines of work that addressed many of these concerns. Since then, experimental and behavioral economics have flourished and have been honored with the Nobel Prizes awarded to Vernon Smith and Daniel Kahneman. Recent advances in psychology are raising new challenges to economic assumptions. On page 1111 of this issue, Marchiori and Warglien (*1*) open a fresh avenue by which we can use models of neural networks to understand how humans learn as they make economic decisions.

Innovations in psychological measurement and theory are dramatically altering our picture of how the mind works, and for economics this means that long-held assumptions about decision-making processes may once again be in tension with accumulating psychological evidence (*2*, *3*). The advent of many types of brain-scanning devices, instrumentation at the level of single neurons, inexpensive hormonal assays, and many other new techniques is propelling a rapid evolution in our understanding of how human beings formulate, carry out, and are changed by their actions.

These developments present a second wave of challenge for economic theory: how to incorporate the essential elements of what psychology is revealing about the interaction among habits, emotions, and choices, while not sacrificing the parsimony and cumulation that are valuable hallmarks of economic theorizing. This tension energizes the newly emergent field of neuroeconomics that bridges the worlds of economics and contemporary psychology (*4*).

Marchiori and Warglien used neural network models that incorporate "regret" to predict the outcomes of games played by humans. In this context, regret refers to the difference between outcomes attained and the best outcomes that might have been attained if the actor had chosen differently. The models' predictions are based not on conventional, forward-looking expectations of gain, the notion so long at the heart of economic theorizing, but instead on the action propensities that develop through a backward-looking learning process that is driven by regret.

This is an important step in the development of a workable new synthesis. Marchiori and Warglien show that a very simple, parameter-free model can do an excellent job of fitting the long-run tendencies of players in 21 different economic gaming experiments. Their approach extends a recent emphasis on obtaining useful approximations to actual choice behavior, which is a different goal from the exact explication of neurological mechanisms sought by psychologists (*5*).

In fact, Marchiori and Warglien tested multiple models, including many that were advantaged by having free parameters that could be estimated from the data. Their simple parameter-free model fit the data on game play better than almost all of the competing models. In particular, it outperformed the classic Nash equilibrium predictions (*6*) for mixed-strategy games—games where a player maximizes payoff by enact-

School of Information, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: mdc@umich.edu

CREDIT: PETER HOEY

ing a sequence of moves, each of which is random but with probabilities that should optimize expected gains. These elegant theoretical results have long been known to provide poor fits to laboratory data from real players.

The success of their model is impressive, but, as they point out, it is only part of the significance of their results. The three best-performing models are (i) a model drawn from the previous literature known as Normalized Fictitious Play (*7*), (ii) the simplest Marchiori-Warglien model, and (iii) a variant of their model with one free parameter. Although they vary in how it is incorporated, all three are regret-driven learning models.

To these results we can add others (*8*) showing that a regret-based model can also account for puzzling empirical patterns in various forms of auction bidding. As Marchiori and Warglien recognize, the superior fit of these regret-based models aligns very nicely with results from recent neuropsychological work, where studies of brain activity during decision-making have pointed increasingly to the crucial role played by brain areas implicated in feelings of regret (*9*). In particular, regret involves constructing a counterfactual image of what the actor would have felt if he or she had acted differently. Psychological subjects with impaired abilities to construct such felt counterfactuals—for example, because of damage to brain areas such as the orbitofrontal cortex—are observed to behave quite differently from normal subjects in choice situations (*10*).

Although the Marchiori-Warglien model gives a central role to regret-driven learning in making its successful predictions, it is important to recognize that the learning in their model is an approximation of the learning dynamics of experimental players, just as the regret is an approximation of the complex feelings of success and failure experienced by the players. The adaptive neural net structure used in their model converges toward its stable propensities for long-run play fairly quickly, generally in fewer iterations than experimental subjects require. Thus, the data on play that it fits least well are the cases where subjects take the longest to stabilize their patterns of play. Other models, typically those based on more traditional reinforcement learning, have been found to do better at tracking the early stages of experimental play (*5*). But versions of those reinforcement models that were included in the present work did less well overall than any of the regret-driven models.

The great virtue of the parsimony and rigor of economic theorizing is exactly that improvements often accumulate rapidly. Thus, the Marchiori-Warglien model may not be the final word in this development of psychologically plausible approximations for economic game play (*11*). However, the authors have made an important contribution with their psychologically grounded insight that regret-driven learning provides the best approximating form to date for some of the most recalcitrant economic laboratory data that we have. And their use of a neural network structure incorporating payoffs has interesting possibilities for generalization to other classes of games. They have made a step that both takes us forward and simultaneously helps to define the road that lies ahead.

References and Notes

1. D. Marchiori, M. Warglien, *Science* **319**, 1111 (2008).
2. C. Camerer, U. Malmendier, in *Economic Institutions and Behavioral Economics: Proceedings of the Yrjö Jahnsson Foundation 50th Anniversary Conference*, P. Diamond, H. Vartiainen, Eds. (Princeton Univ. Press, Princeton, NJ, 2007), pp. 235–280.
3. M. D. Cohen, in *Economic Institutions and Behavioral Economics: Proceedings of the Yrjö Jahnsson Foundation 50th Anniversary Conference*, P. Diamond, H. Vartiainen, Eds. (Princeton Univ. Press, Princeton, NJ, 2007), pp. 281–290.
4. E. Fehr, C. Camerer, *Trends Cogn. Sci.* **11**, 419 (2007).
5. I. Erev, A. E. Roth, R. L. Slonim, G. Barron, *Econ. Theory* **33**, 29 (2007).
6. R. Gibbons, *Game Theory for Applied Economists*, (Princeton Univ. Press, Princeton, NJ, 1992).
7. I. Erev, A. E. Roth, *Am. Econ. Rev.* **88**, 848 (1998).
8. R. Engelbrecht-Wiggans, E. Katok, *Econ. Theory* **33**, 81 (2007).
9. G. Coricelli, R. J. Dolan, A. Sirigu, *Trends Cogn. Sci.* **11**, 258 (2007).
10. A. Bechara, H. Damasio, A. Damasio, *Cereb. Cortex* **10**, 295 (2000).
11. For example, the few cases that the simple model has trouble predicting are ones in which the optimal behavior is fairly close to an unmixed, or pure strategy, suggesting that there are further regularities that may be captured in a subsequent variation on the model.

10.1126/science.1155477

MATERIALS SCIENCE

Bioinspired Structural Materials

Materials scientists are seeking to create synthetic materials based on the mechanical design principles found in biological materials such as seashell nacre.

Christine Ortiz and Mary C. Boyce

The huge diversity of structural biological materials that exist in nature, even within a single species, and the complexity, multifunctionality, and multiscale nature of their structure-property relationships has been studied extensively for decades (*1*). Using materials available in the environment that typically exhibit poor macro-scale mechanical properties (brittle biological ceramics and compliant macromolecules), they can achieve orders-of-magnitude increases in strength and toughness; in many cases, this "mechanical property amplification" occurs in a nonadditive manner that goes beyond the simple composite rule of mixture formulations. Synthetic structural materials that take advantage of the mechanical design principles found in nature could transform many fields; e.g., materials science, mechanical and civil engineering, and aeronautics and astronautics. Here, we highlight a few recent developments in this area and summarize unexplored opportunities for the future.

Bonderer *et al.* [page 1069 of this issue (*2*)] have carried out the deliberate microstructural design of a multilayered alumina platelet-reinforced chitosan nanocomposite, inspired by the inner nacreous layer of many seashells. The alumina platelets possess higher ultimate tensile strength than the aragonite platelets found in nacre; this reflects a general design concept, whereby the weak constituents found in nature are replaced with more advanced synthetic engineered materials, with the goal of producing structural composite materials with mechanical properties that exceed both those of nacre and those of state-of-the-art synthetic materials.

The thickness of the alumina platelets was chosen to be similar to that found in nacre (~200 nanometers). At such sub-micrometer thicknesses, the strength of brittle materials often increases compared with the bulk material because of the decreasing size and probability of flaws (*3*). The average platelet aspect ratio was selected based on a classic composite shear lag load transfer model (*4*); this critical length maximizes the load transferred to the platelet while averting platelet fracture and instead favors a failure mechanism of platelet pullout

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: cortiz@mit.edu

and matrix shear (as observed in nacre). The authors also achieved excellent alignment and dispersion of the platelets in the matrix up to volume fractions of 0.2, using a colloidal-based technique. Tensile mechanical properties were amplified in the composite material compared to the chitosan matrix, and significant plasticity was retained, suggesting pathways for extensive deformation of the matrix at these volume fractions before platelet pullout. The authors have thus created a material that is simultaneously stiff, strong, and tough by via constituents of appropriate length scale and geometry, using mechanical design principles derived from nature, as well as achieving dispersion of the reinforcing component in the matrix.

Podsiadlo and colleagues (*5*) have taken an alternative approach in the design of a montmorillonite clay platelet–poly(vinyl alcohol) matrix nacre-mimetic artificial nanocomposite by focusing on tailoring the chemistry of the platelet-matrix interface to enhance load transfer. Biological composites make use of local chemistry, compositional gradients, macromolecular supramolecular structure, length scale effects, geometry, and other factors to design robust interfaces and interphases that bond together different material phases, even in the presence of water (*6*). In the material prepared by Podsiadlo *et al.* (*4*), the platelets and matrix exhibit extensive epitaxial hydrogen bonding, as well as cyclic cross-linking. Gluteraldehyde treatment created additional covalent acetal bridges between the matrix and platelets, and dramatic increases in stiffness and strength were achieved compared to the matrix material.

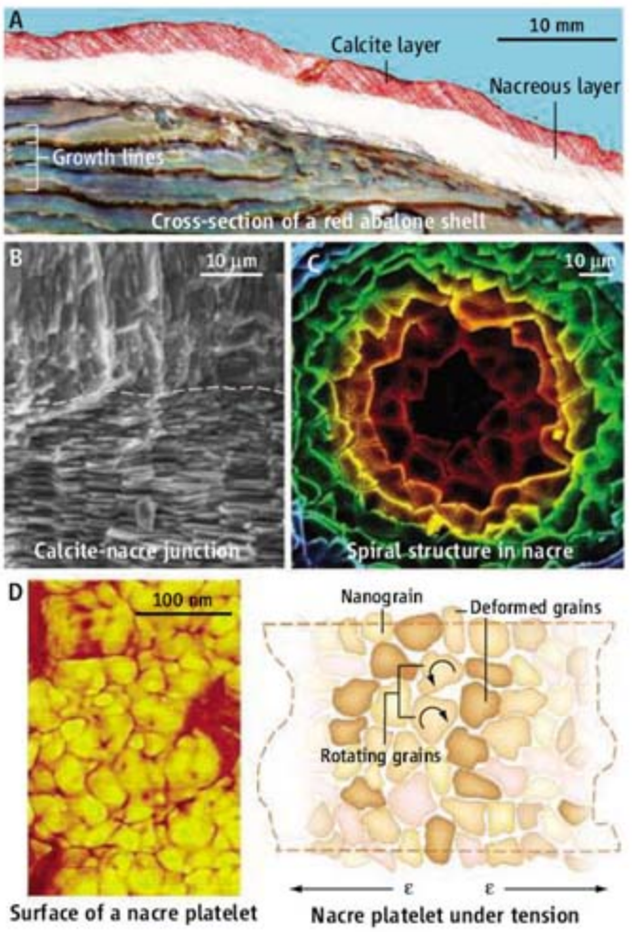
Another example of nacre-mimetic interfacial design between structural elements in a nanocomposite has been reported by Tang *et al.* (*7*). Here, an ionically bonded polyelectrolyte matrix was used to mimic the high toughness extensional "sawtooth" macromolecular elasticity profile due to sacrificial noncovalent bonding observed experimentally in the organic component of nacre (*8*) and modeled theoretically by Qi *et al.* (*9*). This unique interfacial property governs the adhesive matrix behavior, thus controlling and mitigating shear lag load transfer to the aragonite tablets, averting failure of both the tablets and the organic adhesive layers, and increasing ductility and energy dissipation.

Even in the limited context of nacre, many design principles remain to be explored. The ability to access a wide range of volume fractions of the reinforcing component will enable additional design optimization in terms of tailoring stiffness, strength, and toughness. Recent theoretical models (*10*) and electron microscopy studies (*11*) of nacre have shown that additional microstructural features—platelet layer offset, platelet surface waviness, the Voronoi arrangement of platelets in each plane, and screw dislocations leading to large, interconnected layer-to-layer spiral structures (see the figure, panel C) and a tessellated zigzag morphology—may all play a role in the biomechanical functionality.

The smaller and larger length scale design principles of seashells have received less attention. At the smaller length scales, individual nacre platelets are complex organic-inorganic composites with a unique sector structure; each sector possesses nanometer-sized surface domains or asperities (*12, 13*). Individual platelets can exhibit considerable plasticity before fracture upon penetration by an indenter (*12*), and the constituent surface "nanograins" can deform and rotate under an applied tensile load (*13*) (see the figure, panel D). In addition, there is a complex three-dimensional distribution of organic matrix components spatially within the plane parallel to the nacre tablets (*14*).

At the larger length scale of the shell (see the figure, panel A), the multilayered structure, the prismatic calcite outer layer, integrity of the calcite-nacre interface (see the figure, panel B), confinement effects between the layers, structure and property gradation within and between layers, and anisotropy of the layers all work collectively to provide enhanced mechanical performance.

Last, we have yet to fully understand and take advantage of the inherent specificity of natural mechanical design principles. For example, multilayered armored fish scales serve as protection from predatory penetrating impacts (*15*), mussel byssal threads are hysteretic yet resilient to large strain deformation in order to maintain adhesion to rocks in the face of the pounding surf (*16*), and graded layer junctions in teeth resist catastrophic fracture during mastication (*17*). Each of these systems experiences, and has been designed to endure, very different loading conditions in their environment and during their function.



Multiscale mechanical design principles of seashells. (**A**) Multilayered structure of the cross-section of California red abalone shell (*10*). (**B**) Artificially colorized scanning electron microscopy (SEM) image of the calcite-nacre junction region in a *Trochus niloticus* shell (*9*). (**C**) Artificially colorized SEM image of a layer-to-layer spiral in *T. niloticus* nacre. (**D**) (Left) Atomic force microscopy image of surface nanograins on an individual nacre platelet from California red abalone and (right) schematic of nanograin rotation under tension (*13*).

References and Notes

1. S. A. Wainwright, W. D. Biggs, J. D. Currey, J. M. Gosline, *Mechanical Design in Organisms* (Princeton Univ. Press, Princeton, NJ, 1976).
2. L. J. Bonderer, A. R. Studart, L. J. Gauckler, *Science* **319**, 1069 (2008).
3. H. Gao, B. Ji, I. L. Jager, E. Arzt, P. Fratzl, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5597 (2003).
4. A. P. Jackson, J. F. V. Vincent, R. M. Turner, *Proc. R. Soc. London Ser. B.* **234**, 415 (1998).
5. P. Podsiadlo *et al.*, *Science* **318**, 80 (2007).
6. S. Weiner, F. Nudelman, E. Sone, P. Zaslansky, L. Addadi, *Biointerphases* **1**, 12 (2006).
7. Z. Y. Tang, N. A. Kotov, S. Magonov, B. Ozturk, *Nat. Mat.* **2**, 413 (2003).
8. B. L. Smith *et al.*, *Nature* **399**, 761 (1999).
9. H. J. Qi, B. J. F. Bruet, J. S. Palmer, C. Ortiz, M. C. Boyce, in *Mechanics of Biological Tissues*, G. A. Holzapfel, R. W. Ogden, Eds. (Springer, Graz, Austria, 2005), pp. 175–190.
10. F. Barthelat, H. Tang, P. D. Zavattieri, C. M. Li, H. D. Espinosa, *J. Mech. Phys. Solids* **55**, 306 (2007).
11. N. E. Yao, A. Akey, *J. Mat. Res.* **21**, 1939 (2006).
12. B. J. F. Bruet *et al.*, *J. Mat. Res.* **20**, 2400 (2005).
13. X. Li, Z. H. Xu, R. Wang, *NanoLetters* **6**, 2301 (2006).
14. F. Nudelman, B. A. Gotliv, L. Addadi, S. Weiner, *J. Struc. Biol.* **153**, 176 (2006).
15. J. Daget, M. Gayet, F. J. Meunier, J.-Y. Sire, *Fish Fisheries* **2**, 113 (2001).
16. E. Carrington, J. M. Gosline, *Am. Malacological Bull.* **18**, 135 (2004).
17. V. Imbeni, J. J. Kruzic, G. W. Marshall, R. O. Ritchie, *Nat. Mat.* **4**, 229 (2005).

18. The authors would like to acknowledge funding from the U.S. Army (DAAD-19-02-D0002).

10.1126/science.1154295

RETROSPECTIVE

Judah Folkman (1933–2008)

An innovative and inspirational pioneer in the field of angiogenesis research and its medical applications is remembered.

Douglas Hanahan[1] and Robert A. Weinberg[2]

Judah Folkman died on 14 January 2008 en route to a scientific conference on angiogenesis, the field he fathered and championed. His loss has resonated throughout the biomedical research community, where he was celebrated for his boundless creativity, indefatigable optimism, a warm and humble personality, inspirational teaching and advocacy, and empathy as a clinician.

Trained at Harvard Medical School, Folkman had a remarkable intuition for organ biology. In the early 1960s, while serving in the U.S. Navy and stationed in a laboratory investigating blood substitutes, he began studying (and observing limitations to) the growth of explanted tumors, sowing seeds of inquiry that led to his landmark 1971 theoretical paper in which he proposed that tumors depend on the active induction and continuous growth of new blood vessels (angiogenesis) to survive and expand. He envisioned that such tumor-associated angiogenesis is not a passive physiological response, but is regulated by specific factors. He postulated that tumors express angiogenesis-inducing molecules, and that pharmacological inhibitors of tumor angiogenesis could be developed into useful drugs to treat human cancers. He subsequently substantiated early clues that tumors were angiogenesis dependent and developed innovative experimental systems to investigate the biology of angiogenesis.

In 1981, Folkman stepped down from his position as surgeon-in-chief at Children's Hospital, Boston, to focus on angiogenesis research. Since the late 1960s, he had sought to identify the postulated regulatory factors governing tumor angiogenesis, but the task proved challenging. In 1984, his team reported that a growth factor expressed by tumor cells—fibroblast growth factor 2 (basic fibroblast growth factor)—could induce angiogenesis. Their work on purification and bioassays set the stage for others to isolate vascular endothelial growth factor (VEGF), another potent (and more specific) angiogenic protein. VEGF proved to be the same vascular permeability factor discovered earlier by Harold Dvorak, Folkman's colleague at Harvard Medical School.

Folkman's frustration as a clinician with the inadequacies of conventional cancer therapies, and his conviction that blocking angiogenesis would yield therapeutic benefit, led him to focus increasingly on angiogenesis inhibitors, beginning with the discovery in 1975 of an inhibitory activity in cartilage tissue. This was followed by the demonstration in 1982 that two cellular proteins, protamine and platelet factor 4, could inhibit angiogenesis. Later, buoyed by Noël Bouck's 1989–1990 reports of a tumor-derived inhibitor, the protein thrombospondin-1, Folkman's group discovered angiostatin (in 1994) and endostatin (in 1997), cleaved fragments of plasminogen and type XVIII collagen, respectively. These discoveries helped establish a principle that endogenous angiogenesis inhibitors serve as physiological modulators of angiogenesis.

Folkman's contributions and impact on the field are impressive for their breadth. Working with one of us, he demonstrated in 1989 that induction of tumor angiogenesis was not a mere consequence of oncogene-induced hyperproliferation, but rather a discrete event occurring during multistage tumorigenesis, subsequently called "the angiogenic switch." He and his colleagues conceived of a new way to use cytotoxic chemotherapy to inhibit angiogenesis involving a low-dose "metronomic" regimen that is now showing encouraging results in clinical trials. The notion of continuous, rather than bolus, drug delivery had its roots in another Folkman innovation. In the 1960s, he recognized the need for continuous delivery of steroids and performed studies with silicone-tube implants that led to timed-release implant technology for birth-control drugs. These are but examples of a remarkable set of accomplishments.

Folkman's passionate advocacy was not always evenly counterbalanced by the scientist's healthy skepticism. Thus, although his discoveries of endostatin and angiostatin were clearly important scientific achievements, they have not yet proven to be the breakthrough drugs he envisioned. Inspired by a provocative study from Folkman's lab on their effects in tumor-bearing mice, James D. Watson proclaimed in an unguarded moment that "Judah is going to cure cancer in two years." Watson, already a legend for his broad impact on the development of modern biology, saw his prediction emblazoned on the front page of the *New York Times* in May 1998, creating a firestorm. Folkman was suddenly in the public eye, on prime-time TV, and in major news magazines, explaining his theories about angiogenesis and cancer and the potential therapeutic benefits of angiogenesis inhibitors. This publicity produced great expectations to provide these new miracle drugs to patients. Delivering on their promise, however, has not been straightforward. Although angiostatin and endostatin have evident activity as anti-angiogenic agents, neither showed substantial benefit in early-phase clinical trials. Regrettably, the drugs proved expensive to produce and their clinical development was shelved largely for economic considerations, without their therapeutic efficacy being fully tested. Interestingly, however, a formulation of endostatin has been approved for treating lung cancer in China, suggesting its potential. Folkman remained an optimistic advocate of endostatin's prospects, and his vision may yet be realized.

Folkman deeply believed that angiogenesis inhibitors would come to occupy an important place in the treatment of cancer and other diseases with angiogenic etiologies. Recently, three drugs that inhibit VEGF signaling were approved for treating certain cancers, and two more for angiogenic macular degeneration. Their demonstrable but often transitory efficacy represents important proofs of concept. Dozens of other angiogenesis inhibitors are in the drug pipeline, with prospects to further improve clinical responses. The benefits of anti-angiogenic therapy—and more will surely come—will serve as an enduring testament to Folkman's legacy as an innovator and inspiration for the future of angiogenesis research.

10.1126/science.1156080

[1]Department of Biochemistry and Biophysics, University of California–San Francisco, San Francisco, CA 94143, USA. E-mail: dh@biochem.ucsf.edu [2]Whitehead Institute–Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

CREDIT: KRISTEN JOHNSON/CHILDREN'S HOSPITAL, BOSTON, MA

Nuclear Coupling and Polarization in Molecular Transport Junctions: Beyond Tunneling to Function

Michael Galperin,[1] Mark A. Ratner,[1] Abraham Nitzan,[2] Alessandro Troisi[3]

Much current experimental research on transport in molecular junctions focuses on finite voltages, where substantial polarization-induced nonlinearities may result in technologically relevant device-type responses. Because molecules have strong polarization responses to changing charge state or external field, molecules isolated between electrodes can show strongly nonlinear current-voltage responses. For small applied voltages (up to ~0.3 volt), weak interaction between transporting electrons and molecular vibrations provides the basis for inelastic electron tunneling spectroscopy. At higher voltages and for certain time scale regimes, strong coupling effects occur, including Coulomb blockade, negative differential resistance, dynamical switching and switching noise, current hysteresis, heating, and chemical reactions. We discuss a general picture for such phenomena that arise from charging, strong correlation, and polarization (electronic and vibrational) effects in the molecule and at the interface.

Electronic conduction in the single-particle or single-molecule limit can lead to more complex behavior than can be seen in bulk transport systems. Some effects arise from the small number of electrons passing through the molecules; for example, the charging of a molecule by a single electron can give rise to a steplike Coulomb blockade behavior in the current, I, measured as a function of source/drain voltage, V_{sd}. Molecular transport junctions (MJs), in which electrons move through molecules under voltage bias, can show striking nonlinear effects, such as negative differential resistance (drops in I with increasing V) and single-molecule switching accompanied by hysteresis. Many of the most interesting experiments in molecular electronics (ME) require understanding beyond the simplest picture of MJ transport, the limit of coherent tunneling [recently overviewed by two of us (*1*)]. This is the limit of very weak coupling between the transporting charge and the polarization caused by the molecular and environmental vibrational and electronic degrees of freedom that can interact with the moving charge. In this regime, the electron does not charge the molecule (that is, the energy levels do not reorganize) but uses unoccupied energy levels to assist the tunneling process.

In this overview, we explore the more complex, exciting work in MJs focused on the energy regime in which the molecular internal degrees of freedom become involved. This strong vibrational or electronic coupling regime is accessed at higher driving voltage. We can think about these two regimes in terms of how much interaction the tunneling electron is allowed to have with the molecular and environmental vibrations and electronic polarization, that is, the extent of coupling.

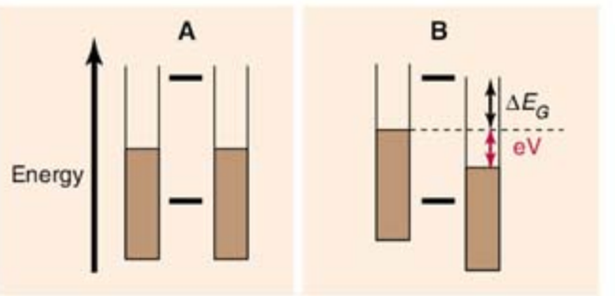


Fig. 1. (**A**) Energy diagram for an unbiased junction, with occupied electronic states in the electrodes in brown and molecular levels as sharp lines. (**B**) A voltage V is applied across the source-drain, but still the molecular levels are not in electronic resonance, because ΔE_G is still large.

Almost 2 decades ago, Yablonovich (*2*) pointed out that, when electron transport changes the charge state of a molecule, the vibration coupling caused by the new potential could be so strong that it could lead to bond breaking, thus obviating any applications in electronic devices. In fact, harnessing electronic changes to access vibrational motion enriches the palette of possible transport behaviors, including unusual switching and memory regimes, and provides an intrinsic scientific challenge as well as possible single-molecule device function. Indeed, MJ science might be seen as now finishing chapter 1 (devoted to establishing methods and investigating the simplest situation of coherent tunneling transport) and entering chapter 2 (examining richer phenomena involving nonlinear transport, charging, vibronic effects, switching, and spectroscopy).

Early theoretical speculations considering the use of single molecules as circuit elements (*3*) and even as extended architectures (*4*), and pioneering work by Kuhn and co-workers (*5*) that measured currents through adsorbed molecular layers on metals and semiconductors, introduced the idea of molecular conduction. New assembly techniques for preparing controlled molecular adlayers allowed fabrication of structures on which conductance could be measured (*6*–*8*). Development of scanning probe microscopy turned ME, and conductance spectroscopy in particular, into scientifically legitimate fields. A major early breakthrough was the invention of scanning tunneling microscopy (STM) (*9*) to examine conductance characteristics of single adsorbed molecules (*10*, *11*). Later developments include mechanical (*12*–*15*) and electrochemical (*16*, *17*) break junctions, allowing the inclusion of a gate electrode.

Diagrams like those in Fig. 1 are often used to discuss MJ transport qualitatively (*18*, *19*). In these one-electron energy-level diagrams, the single-electron energies shown should not be confused with molecular state energies. Molecules have a finite number of electrons and show substantial charge interaction effects both in the purely electronic states and in the vibrational-electronic (vibronic) manifolds. Energies of the frontier molecular orbitals, the highest occupied and the lowest unoccupied molecular orbital (HOMO and LUMO, respectively), can change by more than 1 eV upon charging, even without notable geometry changes (unlike the simpler tunneling case). Such changes arise from electron interactions within the molecule. Coulomb interactions make these one-electron energies dependent on total electron count n, thus,

$$\varepsilon_i = \varepsilon_i(n) \qquad (1)$$

and Fig. 1 must be interpreted quite carefully, because the energies will change upon charging.

Early work on MJs focused on the regime (*1*, *20*) where the two levels corresponding to the HOMO and LUMO in Fig. 1 remain far above and below the relevant Fermi energies of the electrodes. In that case, molecular charging does not take place, transport occurs in a coherent tunneling regime assisted by the molecular levels, and the molecule just acts as a tunneling barrier. This is the Landauer-Imry (L-I) regime (*21*). Although the measurements discussed here are steady-state type, relative time scales, associated with important energy parameters, play a decisive role in characterizing transport regimes. The injection gap, ΔE_G, determines the tunneling traversal time, $\tau_c \sim \hbar/\Delta E_G$, in off-resonance situations (*22*), whereas the spectral density, Γ, is related to the molecule-metal coupling and determines, in the absence of nuclear distortions, the time, $\hbar/\Gamma$, spent by an electron on the molecule before escaping into the metal leads.

Vibronic coupling effects are strong when the time spent by the electron on the bridge is long

[1]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [2]Department of Chemistry, Tel Aviv University, Tel Aviv 69978, Israel. [3]Department of Chemistry and Centre of Scientific Computing, University of Warwick, CV4 7AL Coventry, UK.

enough for the charge to interact with the nuclear dynamics. A criterion for the weak coupling limit is given by $M/\sqrt{\Delta E_G^2 + (\Gamma/2)^2} << 1$, where M is the vibronic coupling. Then the electron is transported too rapidly for it to interact significantly with vibrations, and inelastic tunneling can be treated as a weak perturbation. In the opposite limit of strong coupling, such perturbative treatment breaks down. As applied voltage increases, the effective ΔE_G decreases, so an entirely new experimental regime, involving important vibronic effects, becomes available.

A General Model

Conductance spectroscopy (current/voltage measurement) depends on geometry, energy levels and interaction strengths, and effects of the external thermal bath. Molecular levels have discrete MO energies, whereas the electrodes are macroscopic, with continuous densities of states. An appropriate understanding of transport behavior requires treating the (nonequilibrium) dynamics of the molecule between electrodes with different chemical potentials.

To fully describe transport in MJs, we must understand the electrodes, the molecular vibronic structure and response, possible optical excitations, geometric changes, and thermal effects. To clarify where these arise, we use a crude phenomenological model Hamiltonian to describe the molecule, the leads, and their interaction:

$$H = H_{MOL} + H_{LEAD} + V \quad (2)$$

The molecular Hamiltonian, H_{MOL}, describes the molecular electrons (H_{el}), vibrations (H_{vib}), and their interaction, $H_{el\text{-}vib}$:

$$H_{MOL} = H_{el} + H_{vib} + H_{el\text{-}vib} \quad (3)$$

The electronic part is a sum of free electron and electronic correlation terms:

$$H_{el} = \sum_l \varepsilon_l n_l + (correlations) \quad (4)$$

with n_i the number of electrons in energy level i. The Hamiltonian H_{vib} describes harmonic vibrations on the molecule, and the vibronic mixing ($H_{el\text{-}vib}$) term can be written

$$H_{el\text{-}vib} = \sum_{\ell} M^{\alpha}_{\ell,\ell} Q_{\alpha} n_{\ell} \quad (5)$$

with M_l^{α} as the vibronic coupling strength, n_l as the number of electrons in the level, and Q_α as the dimensionless normal coordinate displacement. The form of Eq. 5 assumes that the most important vibronic interaction arises from a shift in the oscillator equilibrium point that is proportional to the level occupation.

Both the term H_{LEAD} and the molecule-electrode interaction V are represented simply, ignoring interelectronic repulsion and retaining only site energy and neighbor tunneling terms (Hückel-type or tight-binding model).

This minimal description can be supplemented by the interactions between molecular vibrations and the thermal environment, the vibrational anharmonicity, the possible role of photons or excitons, and the electron-electron interactions. Different behavior regimes determined by these parameters can be accessed by changing junction conditions, including applied voltage bias, gate potential, and structural patterns of the molecule.

Most electron transfer (ET) theories for molecular systems, and much MJ modeling, ignore the effect of electronic correlations except in a mean-field approximation. Even within the latter, different molecular charge states will have different energy levels. Ignoring electron correlations can lead to confusion and inaccuracies because they can cause important energy shifts upon charging (see below). The electron-vibration interaction (Eq. 5) gives rise to correlation between electron and nuclear dynamics. These two interactions provide the broad spectrum of behaviors discussed here.

The most powerful description of transport in MJs is given by the nonequilibrium Greens function (NEGF) formalism (*23*, *24*) (although simpler descriptions such as scattering theory may be useful close to equilibrium). In this formalism, terms known as "self-energies" are used to describe the effects on a particular system, in this case the molecule, of interactions (both internal, such as anharmonicities and electron-electron interactions, and external, interactions with the environment in which the molecule is found) that are not included in the system description. These self-energies have direct spectroscopic manifestations, appearing as a shift (real part of the self-energy) and broadening (imaginary part, sometimes called spectral density) of molecular energy levels. In addition, the environmental self-energies determine fluxes of energy and particles into and out of the system. The need to evaluate these self-energies leads to difficult theoretical problems, best analyzed in model systems and generally not yet numerically approachable for realistic junctions.

In the L-I limit, the inequality $\Delta E_G > \Gamma >> k_B T$ (where k_B is Boltzmann's constant and T is temperature) holds, and the effect of vibronic coupling is small. The Landauer formula (Eq. 6 below) is then useful if electronic correlations can be disregarded. For smaller Γ and/or faster thermal relaxation, we approach the so-called Coulomb blockade regime (*14*, *25*–*27*), in which coherence between the charge motion from the left electrode to the molecule and that from the molecule to the right electrode is lost, rate equation approaches are useful [if we think of the molecule as a quantum dot, this is the Coulomb blockade (*27*) limit in mesoscopic structures]. These are also complicated, because the entire Fock space on the molecule (that is, all the possible excitations and populations) may be accessed (*26*). In the inelastic electron tunneling spectroscopy (IETS) regime, the vibronic coupling $H_{el\text{-}vib}$ is a weak perturbation on the elastic scattering picture.

Weak Polarization Effects: Inelastic Tunneling Spectroscopy

In MJs, optical spectroscopy is not yet used as a standard tool for several reasons. Injecting light into a small gap between metallic leads is difficult, and the sample can be as small as a single molecule, implying small signal-to-noise ratio. Nuclear dynamics effects are usually deduced from the vibrational structure in tunneling spectra. This structure is associated with H_{vib} and $H_{el\text{-}vib}$ in Eq. 3; the former represents the molecular vibrations, whereas the latter refers to the charge-state dependence of the equilibrium molecular geometry. IETS appears as a small correction to the transmittance in the L-I regime, arising from small modifications to the transmittance associated with molecular vibrations when the tunneling occurs far from an electronic resonance: Essentially, the moving charge has no time to interact with the nuclear polarization (but it uses a new, inelastic, channel imparting energy to the vibrations).

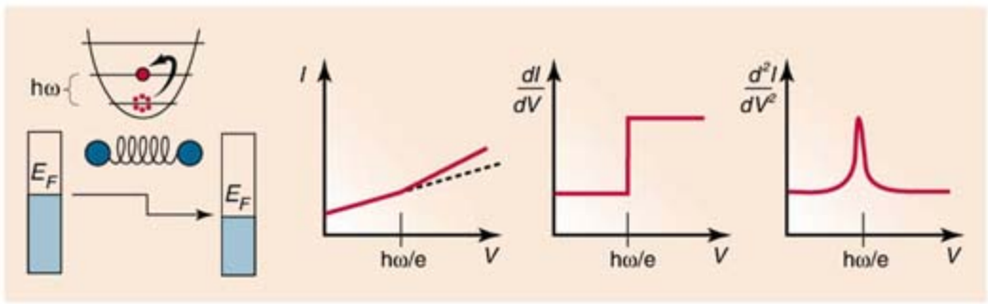


Fig. 2. The expected IETS behavior as the voltage V_{sd} is increased though a vibrational resonance.

Initial observations of IETS in MJ were obtained in STM structures (*28*). Other testbeds, involving molecular adlayers, were then studied, and true single-molecule junctions have been investigated by using IETS (*29*–*31*). As Fig. 2 demonstrates, for bias energy eV_{sd} less than the vibrational energy $\hbar\omega$, the tunneling is elastic. Above a threshold where $\hbar\omega = eV_{sd}$, the transporting electron can deposit $\hbar\omega$ on the molecule. A second inelastic channel could then open, in addition to the elastic one. The cross section for such an excitation is very small because the electron traversal time ($\sim\hbar/\Delta E_G$) is much smaller than the oscillator period. One expects such a

threshold for each normal coordinate of the molecule, unless propensity rules limit such excitation (see below).

The expectation for what one would see on the basis of a simple channel addition model is shown in Fig. 2. The change caused by the opening of the vibrational resonance channel at $eV_{sd} = \hbar\omega$ are usually not visible because the cross section is so small. Such changes can, however, be observed as reproducible features in the second derivative d^2I/dV_{sd}^2 plotted against V_{sd}. These features are usually peaks, indicating increase in transmission upon opening of a vibrational channel, but dips, that is, negative peaks, resulting from interference between the elastic and inelastic transmission are sometimes observed as well. A plot of d^2I/dV_{sd}^2 versus V_{sd} is referred to as the IETS spectrum. Experimental measurements made in a crosswire junction at 9 K are shown in Fig. 3.

Elegant theoretical approaches to IETS (*32*–*37*) include some that deal with a correct analysis of the line shapes (*38*, *39*) rather than simply assuming that the elastic and inelastic channels are additive (the basis for Fig. 2). The weakness of the spectrum suggests use of perturbation theory, which has reproduced experiments well (*40*), and allows data interpretation. The L-I coherent conductance, g, is

$$g = 2e^2/h\, Tr[\Gamma^R G \Gamma^L G^+] \qquad (6)$$

$$(G)^{-1} = E - H + (i/2)\Gamma \qquad (7)$$

where Γ^R and Γ^L are spectral densities in the right and left electrodes, respectively, whereas G is the (retarded) Green's function, with H the system Hamiltonian. Using a normal-coordinate expansion and standard DFT calculations, we can compute the conductance of the inelastic channel (and the IETS spectrum). Propensity rules for IETS can be also be derived, but unlike selection rules for Raman or infrared spectroscopy these rules do not arise from the symmetry properties of the light-matter interaction. Rather, the rules follow from invariance properties of the Green's function matrix, and some arguments about intensities are expected along particular pathways (*41*–*43*). They are helpful in assigning spectra, discerning chemical reactions in the junction, obtaining a pathway description for tunneling, and finding the geometry in MJs.

Because IETS measurements are sensitive to molecular geometry and charge (*42*, *43*), they might help understand processes outside the L-I regime, in particular when charge on the molecule changes with applied voltage (where strong electron-vibration coupling can also lead to new functional properties).

Strong Polarization Effects

IETS is the most useful, reliable characterization tool available today for MJs. Other important phenomena are associated (*44*) with polarization interactions in junctions:

1) Resonance inelastic electron tunneling spectroscopy (RIETS) is related to the standard IETS just as resonance Raman scattering is related to Raman scattering. Whereas the normal, weak coupling processes provide information about the ground molecular state, their resonance counterparts are sensitive to vibrational structure and motion in the excited state (for Raman scattering) or the intermediate molecular ion (in MJs).

2) Strong vibronic and polarization interactions are implicated in charging phenomena. Charging coupled to nuclear motion can lead to nonlinear phenomena, such as bistability, hysteresis, switching, and negative differential resistance (NDR), that pertain to junction functionality. The coupling of change in electronic charge distribution with response of the nuclear configuration dominates molecular ET. The difference in MJs is that the initial and final charging states are nonequilibrium aspects of the current-carrying junction.

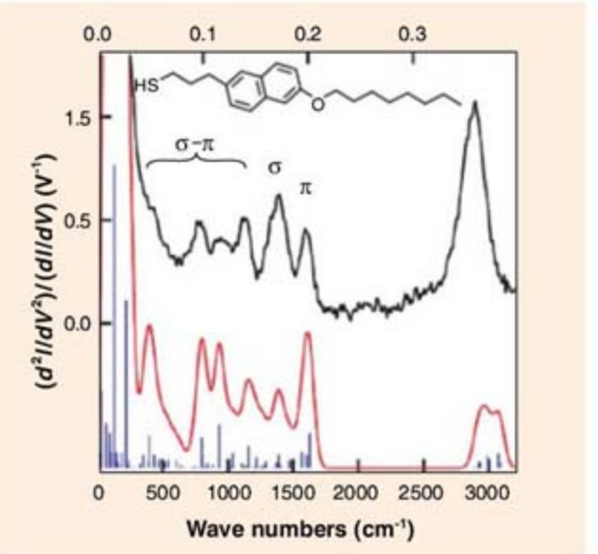


Fig. 3. Experimental (black) and computed (red) IETS spectra for the molecule in the inset, with modes specified as coupled to σ or π type tunneling. Blue lines indicate the computed frequency and IETS intensity of the individual modes [from (*43*)].

3) Strong vibronic interactions can lead to strong junction heating. Generated heat can be carried away by both phonons and electrons. Heat generation and dissipation on the molecular scale are therefore important aspects of overall junction operation.

4) Affecting molecular configuration changes and reactivity, by the combined effects of strong field and electronic current in a MJ, is another potentially important application of such junctions.

Such issues constitute some of the frontline problems in studies of MJs. Here, we briefly overview their present status.

RIETS. Under electronic resonance tunneling conditions, the injection energy measured relative to vacuum is similar to the molecular ionization energy (for hole injection) or electron affinity (for electron injection), often expressed in simplified discussions as HOMO and LUMO energies. A transient intermediate molecular ion forms and may be stabilized by electronic or nuclear polarization of the environment (image or polaron formation). The vibrational structure of the transient polaron may then appear in the inelastic signal as satellite peaks (sidebands) in the conductance-voltage plot near the conduction threshold. For this structure to be resolved, another inequality, $\omega_0 > \Gamma/2$ (where ω_0 is the relevant vibrational frequency), must be satisfied between the system time and energy scales. These satellite peaks are expected (and seen) in the conduction spectrum rather than in d^2I/dV_{SD}^2 (as in IETS).

Vibronic interactions and charging phenomena. Interest in MJs stems partly from functionalities such as rectification, switching, and addressable memory. Change of redox state has been offered as an underlying cause for such behaviors, supported by experimental observations (*45*–*49*). Mechanisms and dynamics of such transitions are still under discussion. It is clear that nuclear motion and electron-phonon coupling are strongly involved. As in ET reactions, nuclear motion is not merely a consequence of redox state change. Rather, this transition is a cooperative vibronic process. In a mean field model (*50*), such multistability results from positive feedback: Formation of the transient molecular ion polarizes the electronic and the nuclear environments and further drives and stabilizes ion formation. The resulting steady state in the junction leads to a self-consistency condition in which the population of the MO and its energy are interdependent: The orbital energy is shifted by polaron formation by an amount that depends on its electronic population n. The self-consistency condition allows for multistability properties.

Whether this multistability can lead to hysteresis behavior and memory effects as suggested in (*50*) or to intermittent noise associated with transitions between two locally stable states as discussed in (*51*, *52*) is an issue of relative time scales: the observation time versus the rate of transitions between locally stable states, as is evident in the experimental results (*53*) displayed in Fig. 4. This polarization mechanism may also cause NDR phenomena (*48*–*50*), again depending on relative time scales (*54*, *55*). Experimental observations have not yet fully converged. For example, the predicted dependence on gate potential (*50*) is a central observation in (*48*, *49*) but is far less evident in (*56*). The model of (*50*) suggests a vibronic and electronic polarization origin (*57*) of observed

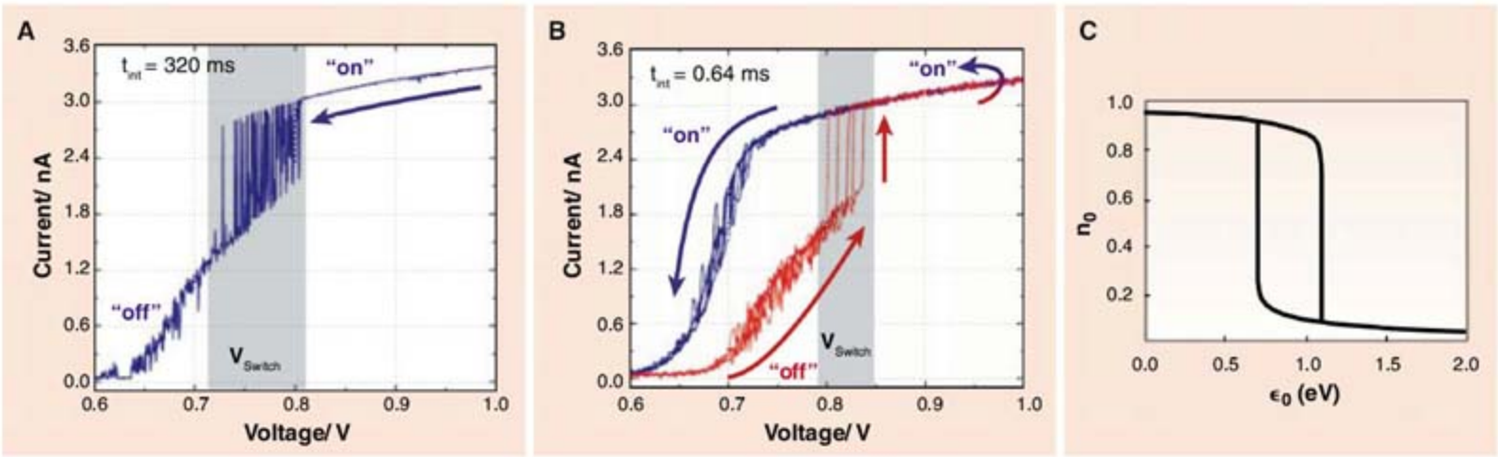


Fig. 4. Dynamics of I/V_{SD} switching response of bipyridyl-dinitro oligophenylene ethynylene dithiol (BPDN-DT) molecules between gold contacts (*53*). (**A**) The voltage is changed relatively slowly, and bistability gives rise to telegraphic switching noise. When voltage changes more rapidly (**B**), bistability is manifested by hysteretic behavior. (**C**) The locally stable solutions of the equilibrium self-consistent relation (*50*) between the energy and population of the molecular level of Fig. 1. Multistability caused by polarization of the environment by the molecular charge suggests that hysteresis, switching behavior, and telegraph-type fluctuations may be observed.

charging transitions in redox MJs but is too simple to account quantitatively for the *I*/*V* behavior; a simple model comparison is shown in Fig. 5. Other mechanisms, for example, modulation of the molecule-lead bonding (*56*, *58*), may also lead to switching and bistability phenomena. These switching mechanisms rely on molecular polarization, in sharp contrast to traditional semiconductor voltage-gated switching.

Heating and heat conduction. The combination of small molecular heat capacity and inefficient heat transfer away from the junction (*2*) might affect the stability and integrity of MJs operating outside the L-I regime. New work aims to assess both heat conduction and temperature rise in a current-carrying junction. When heat transport is dominated by phonons, one can derive (in the harmonic approximation) a Landauer-type equation for the heat current. Such expressions were used to analyze heat conduction properties of alkanes, predicting interesting dependence on molecular chain length (*59*).

The temperature rise in MJs is determined by the balance between heat dissipation and conduction rates. General theories address this balance, but application to realistic junction models is prohibitively complex. Simple models (*59*) show two crossover transitions. The first occurs when the bias matches the frequency ($eV_{sd} > \hbar\omega_0$) and vibrational excitation becomes possible. The second occurs when the voltage permits electronic resonance. Observations pertaining to thermal transport properties of molecular junctions have been recently published (*60*–*63*).

Current-induced reactions (*64*, *65*). Conformational changes and chemical reactions can originate from forces exerted on the molecule, including short-range forces exerted by a tip used to push atoms, long-range electrostatic forces from the imposed potential bias, and forces associated with the transporting current. The ultimate result of affecting chemical change depends on the balance between pumping energy into molecular bonds and dissipation of this excess energy.

New Horizons

Although IETS has been under study for nearly a decade and the above discussion may be considered a report on work in progress, we now reflect on several (unrelated) issues that we expect to become prominent in MJ research.

Understanding electronic correlations. Such phenomena as Coulomb blockade and Kondo resonances arise from strong electronic correlations that are beyond the applicability of mean-field models. Even image effects, reflecting correlations between tunneling electrons and electrons in the metal leads, are difficult to describe properly (*66*). Computational efforts toward improving this have been reported (*67*). Kondo-assisted injection might advance organic photovoltaics, and Coulomb blockade structures could lead to single-molecule memories.

Noise in molecular junctions. Noise has been extensively studied in mesoscopic conductors (*68*), and applications in MJs have been discussed (*69*–*74*). Noise measurements can be useful, for example, for determining the number of channels that contribute to transport and for estimating junction coupling parameters such as the electron/vibration coupling strength and the relative sizes of Γ, the inverse electron lifetime, at the two electrodes (*72*). Different aspects of the effect of vibronic coupling on the current noise have been studied [see (*75*) and references therein]. Further progress awaits a systematic experimental study of this issue.

Illumination effects and junction spectroscopy. The slow introduction of junction spectroscopy in ME can be understood by the unfavorable ge-

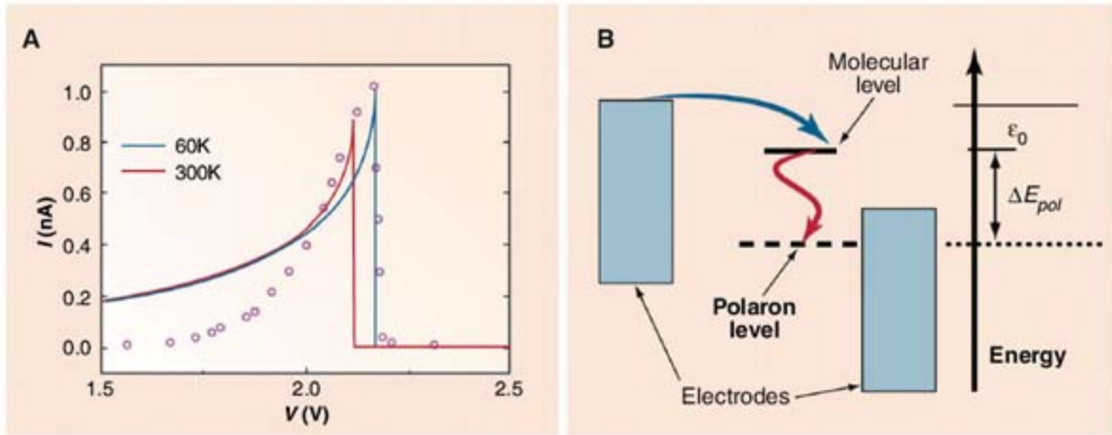


Fig. 5. (**A**) Observed (*48*) (dots) and calculated (in the simple polarization model, solid lines) NDR behavior. (**B**) How NDR can occur. Raising the voltage permits charging of the LUMO level, that subsequently is reduced in energy by vibronic and electronic polarization. When the total polarization energy, ΔE_{pol}, is large enough, the level energy drops below the leads chemical potentials, and (at low *T*) the current drops sharply. At higher *T*, the line broadens and shifts to lower voltage, because of thermal broadening of E_F.

ometry, where the molecule (or molecular layer) is located in a gap of molecular dimension between two essentially macroscopic leads. Recent reports using light to switch MJs (*76*–*82*) indicate that this barrier can be overcome. Indeed, light emission from biased STM junctions can be detected, and although most of the observations are associated with emission by metal plasmons there are some observations of molecular emission. Junction spectroscopy may thus become an important new avenue of research in ME, and the essential proximity of metal and/or semiconductor structures suggests a possible connection with plasmonics.

Accurate computation of transport behavior. As mentioned several times, straightforward quantitative computational methods (at the mean-field level) are available in the L-I regime of elastic tunneling and for IETS (Fig. 3). When polarization, photonic, or correlation effects are present, only model calculations are available today (such as those yielding the computational results of Figs. 4 and 5). A challenging theoretical problem is extension of these formal techniques to allow accurate numerical predictions.

The theoretical constructs and modeling discussed here suggests exciting experimental challenges, including characterizing single-molecule switching and its dynamics, noise and heating measurements, higher-order transport phenomena such as Seebeck coefficients (*83*) or thermopower, and spectroscopy (Raman or optical) in MJs. Many of these arise from the polarization behavior characteristic of molecules (much more than in most semiconductors). In particular, single-molecule switching (at the ~1-nm scale, well beyond current integrated circuit scaling) can occur via vibronic mechanisms that are totally different from the field-effect transistor switching. In sharp contrast to the L-I regime [in which stochastic switching arises from geometry modulations at the contacts (*84*, *85*)], in these vibronic situations switching, charging, and polarization mechanisms provide hugely varying conductance properties among different molecules.

The vibronic and electronic polarization effects discussed quite early (*2*) as possible showstoppers for ME actually can provide single-molecule switching (Fig. 4), with use of only two electrodes. This is based on polarization and oxidation state change, as opposed to the voltage modification in traditional complementary metal-oxide semiconductor circuitry [where the switching is caused by voltage on a third (gate) electrode]. The molecule switches by pulling a switch! Some of these polarization and correlation effects are only now starting to be clarified and promise to lead to answers to device applications and to important new questions.

References and Notes

1. A. Nitzan, M. A. Ratner, *Science* **300**, 1384 (2003).
2. E. Yablonovitch, *Science* **246**, 347 (1989).
3. A. Aviram, M. A. Ratner, *Chem. Phys. Lett.* **29**, 277 (1974).
4. F. L. Carter, *J. Vac. Sci. Technol. B* **1**, 959 (1983).
5. H. Kuhn, D. Mobius, *Angew. Chem. Int. Ed. Engl.* **10**, 620 (1971).
6. D. L. Allara, R. G. Nuzzo, *Langmuir* **1**, 45 (1985).
7. J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, G. M. Whitesides, *Chem. Rev.* **105**, 1103 (2005).
8. Y. Selzer, D. L. Allara, *Annu. Rev. Phys. Chem.* **57**, 593 (2006).
9. G. Binning, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **49**, 57 (1982).
10. L. A. Bumm *et al.*, *Science* **271**, 1705 (1996).
11. N. P. Guisinger, M. E. Greene, R. Basu, A. S. Baluch, M. C. Hersam, *Nano Lett.* **4**, 55 (2004).
12. J. Chen, M. A. Reed, A. M. Rawlett, J. M. Tour, *Science* **286**, 1550 (1999).
13. R. H. M. Smit *et al.*, *Nature* **419**, 906 (2002).
14. J. Park *et al.*, *Nature* **417**, 722 (2002).
15. M. Mayor, H. B. Weber, *Angew. Chem. Int. Ed.* **43**, 2882 (2004).
16. B. Xu, N. J. Tao, *Science* **301**, 1221 (2003).
17. W. Haiss *et al.*, *Nat. Mater.* **5**, 995 (2006).
18. G. Cuniberti, G. Fagas, K. E. Richter, Eds., *Introducing Molecular Electronics: A Brief Overview*, Springer Lecture Notes in Physics (Springer, Berlin, 2005).
19. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge Univ. Press, Cambridge, 2005).
20. N. S. Hush, *Ann. NY Acad. Sci.* **1006** (2003).
21. M. Buttiker, Y. Imry, R. Landauer, S. Pinhas, *Phys. Rev. B* **31**, 6207 (1985).
22. A. Nitzan, J. Jortner, J. Wilkie, A. L. Burin, M. A. Ratner, *J. Phys. Chem. B* **104**, 5661 (2000).
23. L. P. Kadanoff, G. Baym, *Quantum Statistical Mechanics* (Benjamin, Reading, MA, 1962).
24. S. Datta, *Electric Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge, 1995).
25. M. Poot *et al.*, *Nano Lett.* **6**, 1031 (2006).
26. B. Muralidharan, A. W. Ghosh, S. Datta, *Phys. Rev. B* **73**, 155410 (2006).
27. K. K. Likharev, *Proc. IEEE* **87**, 606 (1999).
28. B. C. Stipe, M. A. Rezaei, W. Ho, *Science* **280**, 1732 (1998).
29. J. G. Kushmerick *et al.*, *Nano Lett.* **4**, 639 (2004).
30. W. Y. Wang, T. Lee, I. Kretzschmar, M. A. Reed, *Nano Lett.* **4**, 643 (2004).
31. L. T. Cai *et al.*, *Nano Lett.* **5**, 2365 (2005).
32. A. Troisi, M. A. Ratner, A. Nitzan, *J. Chem. Phys.* **118**, 6072 (2003).
33. N. Sergueev, D. Roubtsov, H. Guo, *Phys. Rev. Lett.* **95**, 146803 (2005).
34. Y. C. Chen, M. Zwolak, M. Di Ventra, *Nano Lett.* **5**, 621 (2005).
35. G. C. Solomon *et al.*, *J. Chem. Phys.* **124**, 094704 (2006).
36. J. Jiang, M. Kula, W. Lu, Y. Luo, *Nano Lett.* **5**, 1551 (2005).
37. M. Paulsson, T. Frederiksen, M. Brandbyge, *Nano Lett.* **6**, 258 (2006).
38. M. Galperin, M. A. Ratner, A. Nitzan, *J. Chem. Phys.* **121**, 11965 (2004).
39. M. Galperin, M. A. Ratner, A. Nitzan, *Nano Lett.* **4**, 1605 (2004).
40. A. Troisi, M. A. Ratner, *Phys. Rev. B* **72**, 033408 (2005).
41. A. Troisi, M. A. Ratner, *Nano Lett.* **6**, 1784 (2006).
42. D. P. Long *et al.*, *Nat. Mater.* **5**, 901 (2006).
43. A. Troisi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14255 (2007).
44. M. Galperin, M. A. Ratner, A. Nitzan, *J. Phys. Cond. Matter* **19**, 103201 (2007).
45. E. A. Osorio *et al.*, *Adv. Mater.* **19**, 281 (2007).
46. X. Y. Xiao *et al.*, *Chem. Phys.* **326**, 138 (2006).
47. J. He, Q. Fu, S. Lindsay, J. W. Ciszek, J. M. Tour, *J. Am. Chem. Soc.* **128**, 14828 (2006).
48. J. Chen *et al.*, *Appl. Phys. Lett.* **77**, 1224 (2000).
49. J. D. Le *et al.*, *Appl. Phys. Lett.* **83**, 5518 (2003).
50. M. Galperin, M. A. Ratner, A. Nitzan, *Nano Lett.* **5**, 125 (2005).
51. A. Mitra, I. Aleiner, A. J. Millis, *Phys. Rev. Lett.* **94**, 076404 (2005).
52. D. Mozyrsky, M. B. Hastings, I. Martin, *Phys. Rev. B* **73**, 035104 (2006).
53. E. Lortscher, J. W. Ciszek, J. Tour, H. Riel, *Small* **2**, 973 (2006).
54. R. A. Kiehl, J. D. Le, P. Candra, R. C. Hoye, T. R. Hoye, *Appl. Phys. Lett.* **88**, 172102 (2006).
55. A. La Magna, I. Deretzis, *Phys. Rev. Lett.* **99**, 136404 (2007).
56. Z. K. Keane, J. W. Ciszek, J. M. Tour, D. Natelson, *Nano Lett.* **6**, 1518 (2006).
57. J. Jortner, *J. Chem. Phys.* **64**, 4860 (1976).
58. A. V. Danilov, S. E. Kubatkin, S. G. Kafanov, K. Flensberg, T. Bjornholm, *Nano Lett.* **6**, 2184 (2006).
59. D. Schwarzer, P. Kutne, C. Schroder, J. Troe, *J. Chem. Phys.* **121**, 1754 (2004).
60. R. Y. Wang, R. A. Segalman, A. Majumdar, *Appl. Phys. Lett.* **89**, 173113 (2006).
61. Z. Wang *et al.*, *Science* **317**, 787 (2007).
62. Z. F. Huang, B. Q. Xu, Y. C. Chen, M. Di Ventra, N. J. Tao, *Nano Lett.* **6**, 1240 (2006).
63. D. Segal, A. Nitzan, P. Hanggi, *J. Chem. Phys.* **119**, 6840 (2003).
64. T. Seideman, *J. Phys. Cond. Matter* **15**, R521 (2003).
65. N. Lorente, R. Rurali, H. Tang, *J. Phys. Cond. Matter* **17**, S1049 (2005).
66. F. Zahid *et al.*, *J. Chem. Phys.* **123**, 064707 (2005).
67. G. Fagas, P. Delaney, J. C. Greer, *Phys. Rev. B* **73**, 241314 (2006).
68. Y. M. Blanter, M. Buttiker, *Phys. Rep.* **336**, 1 (2000).
69. S. Dallakyan, S. Mazumdar, *Appl. Phys. Lett.* **82**, 2488 (2003).
70. J. Lagerqvist, Y. C. Chen, M. Di Ventra, *Nanotechnology* **15**, S459 (2004).
71. K. Walczak, *Phys. Status Solidi B* **241**, 2555 (2004).
72. G. C. Solomon *et al.*, *Nano Lett.* **6**, 2431 (2006).
73. A. Thielmann, M. H. Hettler, J. Konig, G. Schon, *Phys. Rev. B* **71**, 045341 (2005).
74. D. Djukic, J. M. van Ruitenbeek, *Nano Lett.* **6**, 789 (2006).
75. M. Galperin, M. A. Ratner, A. Nitzan, *Phys. Rev. B* **74**, 075326 (2006).
76. D. Dulic *et al.*, *Phys. Rev. Lett.* **91**, 207402 (2003).
77. Y. Wakayama *et al.*, *Appl. Phys. Lett.* **85**, 329 (2004).
78. S. Yasutomi, T. Morita, Y. Imanishi, S. Kimura, *Science* **304**, 1944 (2004).
79. J. He *et al.*, *Nanotechnology* **16**, 695 (2005).
80. T. Kudernac, S. J. van der Molen, B. J. van Wees, B. L. Feringa, *Chem. Commun.* 3597 (2006).
81. N. Katsonis *et al.*, *Adv. Mater.* **18**, 1397 (2006).
82. R. Beckman *et al.*, *Far. Disc* **131**, 9 (2006).
83. P. Reddy *et al.*, *Science* **315**, 1568 (2007); published online 14 February 2007 (10.1126/science.1137149).
84. A. M. Moore *et al.*, *J. Am. Chem. Soc.* **128**, 1959 (2006).
85. H. Basch, R. Cohen, M. A. Ratner, *Nano Lett.* **5**, 1668 (2005).

86. A.N. thanks the Israel Science Foundation, the U.S.–Israel Binational Science Foundation, and the German Israel Fund. M.A.R. is grateful to the NSF (Chemistry, Materials Research Science and Engineering Centers, and International), Defense Advanced Research Projects Agency MolApps, NASA, and U.S. Department of Defense Multiple University Research Initiative. A.T. is grateful to the Royal Society (UK) for support. We thank P. Szuromi for very helpful suggestions.

10.1126/science.1146556

Juvenile Hormone Regulates Butterfly Larval Pattern Switches

Ryo Futahashi[1,2] and Haruhiko Fujiwara[1]*

The spectacular diversity of adult insect color patterns can also extend to differences among sequential larval instars within some species. The swallowtail butterfly, *Papilio xuthus*, represents such an example: Young caterpillars (from the first to the fourth instars) are mimics of bird droppings, whereas the larger, final larval instar (the fifth) has a completely different pattern that is well camouflaged among the leaves of the host plant (Fig. 1A). Here, we show that this developmental switch is regulated by juvenile hormone (JH), which is known to regulate the overall black or green forms caused by differing environmental conditions in some larvae of Lepidoptera and some adults of Orthoptera (*1*, *2*). We applied a JH analog (JHA) to the dorsal surfaces of fourth instar larvae and observed the color pattern of the fifth instar larvae. Treatment at the beginning of the fourth instar stage resulted in a high proportion (67%) of treated larvae reproducing the mimetic pattern instead of the usual cryptic one (Fig. 1B and table S1). No effects were observed when JHA was applied to larvae 20 hours after the appearance of the fourth instar (there are about 96 hours in the total fourth instar stage), indicating that there is a JH-sensitive period (circa 0 to 20 hours after the third ecdysis, Fig. 1B, gray box). A 100-fold lower dose of JHA (50 ng of fenoxycarb) had no effect (table S1). We analyzed JH titers in *P. xuthus* by liquid chromatography–mass spectrometry and found that JH titers decrease during the fourth larval instar (Fig. 1B and fig. S1A). These results suggest that JH regulates the larval pattern switch.

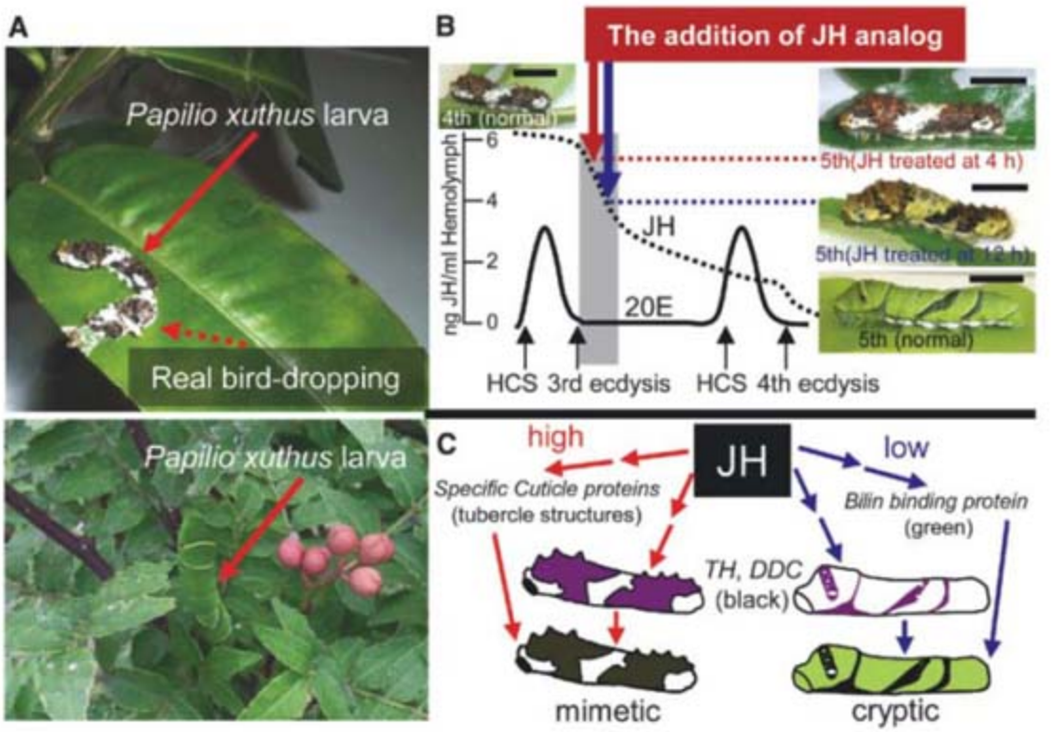


Fig. 1. (**A**) Mimetic and cryptic pattern of *P. xuthus* larva. (**B**) Model of JH titer and its determination of larval body pattern (table S1). Gray box indicates a JH-sensitive period. HCS indicates head capsule slippage, the clear sign of molting period. 20E, 20-hydroxyecdysone. Scale bars indicate 1 mm. (**C**) Model of JH effect on expression of genes associated with mimetic and cryptic pattern (fig. S1).

The major differences between the mimetic and the cryptic larval color patterns include the green coloration of the cryptic pattern, specific tubercle structures (arrowheads in fig. S1D) as components of the mimetic pattern, and the distribution of black pigment. We examined the effects of JHA treatment on gene expression associated with these three differences. By using cDNA subtraction methods, we cloned the hard cuticle protein genes (*HCP1* and *HCP2*) associated with the specific tubercle structures and the bilin-binding protein gene (*BBP*) that is only expressed at the final molt [supporting online material (SOM) text]. JH treatment induced the expression of tubercle-associated cuticle protein genes and inhibited *BBP* expression at the fourth molt, as is the case for the normal third molt (fig. S1, B and C). Larval black patterning is regulated by co-localization of melanin synthesis genes tyrosine hydroxylase (TH) and dopa decarboxylase (DDC) (*3*). The spatial expression patterns of these genes are changed to the mimetic pattern in the JH-treated specimen (fig. S1D). These results suggest that JH regulates the stage-specific gene expression pattern in *P. xuthus* (Fig. 1C) that is required to modulate the larval pattern from mimetic to cryptic.

Our results suggest that a high titer of JH induces the expression of genes associated with the mimetic pattern and that a decrease in JH titer causes a switch to the cryptic pattern. In addition to overall coloration, JH also regulates exoskeletal structures and pigment distribution at specific markings. Because progressive changes of green and black coloration and exoskeletal structures are frequently found in lepidopteran larvae, our findings imply that JH regulation on progressive larval pattern switches may commonly exist.

References and Notes

1. H. F. Nijhout, *Bioscience* **49**, 181 (1999).
2. Y. Suzuki, H. F. Nijhout, *Science* **311**, 650 (2006).
3. R. Futahashi, H. Fujiwara, *Dev. Genes Evol.* **215**, 519 (2005).

4. We would like to thank P. M. Brakefield for critical reading and many helpful comments on the manuscript. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and Research Fellowship of Japan Society for the Promotion of Science for Young Scientists. The nucleotide sequences of *HCP1*, *HCP2*, and *BBP* have been deposited with GenBank (accession numbers AB264673, AB264675, and AB264632, respectively).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1061/DC1
Materials and Methods
SOM Text
Fig. S1
Table S1
References and Notes

27 August 2007; accepted 28 November 2007
10.1126/science.1149786

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bioscience Building 501, Kashiwa, Chiba 277-8562, Japan. [2]National Institute of Agrobiological Sciences, 1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan.

*To whom correspondence should be addressed. E-mail: haruh@k.u-tokyo.ac.jp

A Photon Turnstile Dynamically Regulated by One Atom

Barak Dayan,[1] A. S. Parkins,[1]* Takao Aoki,[1]† E. P. Ostby,[2] K. J. Vahala,[2] H. J. Kimble[1]‡

Beyond traditional nonlinear optics with large numbers of atoms and photons, qualitatively new phenomena arise in a quantum regime of strong interactions between single atoms and photons. By using a microscopic optical resonator, we achieved such interactions and demonstrated a robust, efficient mechanism for the regulated transport of photons one by one. With critical coupling of the input light, a single atom within the resonator dynamically controls the cavity output conditioned on the photon number at the input, thereby functioning as a photon turnstile. We verified the transformation from a Poissonian to a sub-Poissonian photon stream by photon counting measurements of the input and output fields. The results have applications in quantum information science, including for controlled interactions of single light quanta and for scalable quantum processing on atom chips.

The charge and spin degrees of freedom of massive particles have relatively large long-range interactions, which enable nonlinear coupling between pairs of atoms, ions, electrons, and diverse quasi-particles. An early example for electrons is the observation of Coulomb blockade, in which charge transport through small metallic and semiconductor devices occurs electron by electron (*1*–*3*). Among diverse applications for such strong interactions, quantum information science (QIS) relies on large coherent couplings for the implementation of quantum computation, communication, and metrology (*4*).

Whereas electrons interact directly via Coulomb repulsion, photons have vanishingly small cross sections for direct coupling. Instead, photon interactions must be mediated by a material system. Even then, typical materials produce photon-photon coupling rates that are orders of magnitude too small for nontrivial dynamics with individual photon pairs. The leading exception to this state of affairs is cavity quantum electrodynamics (cQED), where strong interactions between light and matter at the single-photon level have enabled a wide set of scientific advances, including single atoms coupled to optical and microwave resonators (*5*–*8*), quantum dots paired with micropillars and photonic bandgap cavities (*9*, *10*), and Cooper pairs interacting with superconducting resonators (*11*, *12*).

Indeed, in analogy with Coulomb blockade for electrons (*1*–*3*), photon-photon interactions in a nonlinear optical cavity were proposed to realize photon blockade (*13*), for which a first photon within an optical system blocks the transmission of a second photon, leading to an orderly output of photons one by one. The initial observation of photon blockade (*14*) for an atomic system used a Fabry-Perot cavity containing one atom strongly coupled to the cavity field, for which the underlying blockade mechanism was the quantum anharmonicity of the ladder of energy levels for the composite atom-cavity system (*15*). Resonant absorption of a first photon to reach the lowest energy eigenstate blocks the absorption of a second photon, because transitions to higher-lying eigenstates are detuned from resonance by the strong interaction of atom and photon. Photon blockade has also been investigated within the context of a photon turnstile (*16*) and has been realized for diverse semiconductor systems (*10*) because of anharmonicity of the photoluminescence spectrum.

In these examples (*10*, *14*–*16*), as well as the original work in (*13*) and extensions thereof (*17*–*19*), photon blockade arises as a struc-

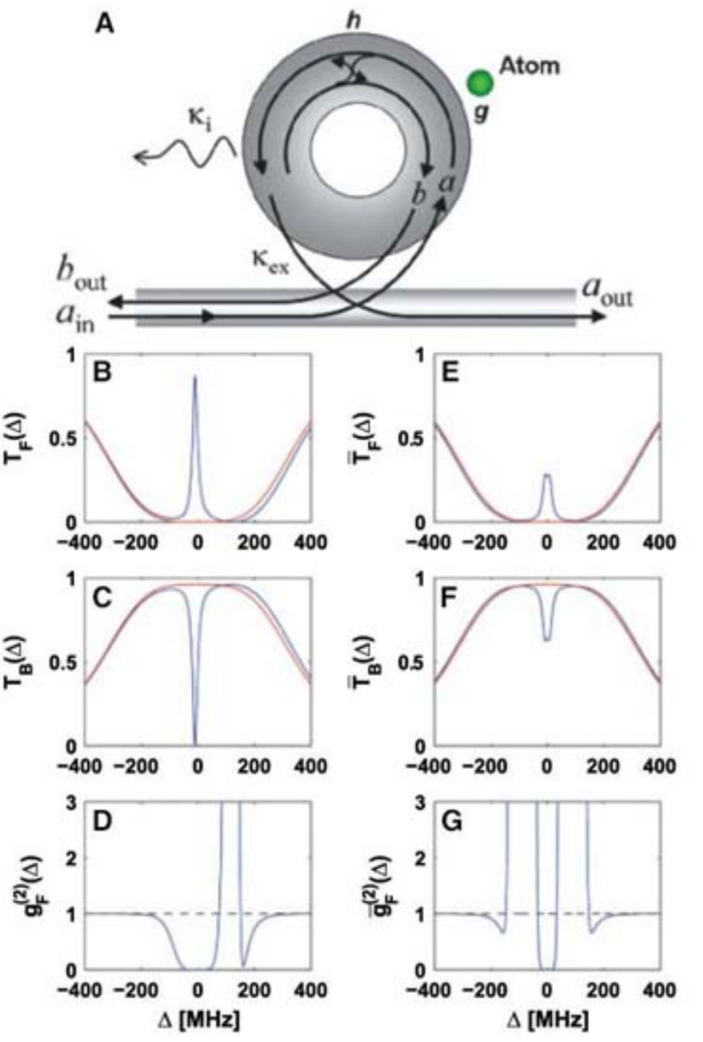


Fig. 1. (**A**) Schematic of microtoroidal resonator and fiber coupler (*20*, *22*). An atom interacts at rate g with the evanescent fields of two internal modes (a and b), which are coupled by scattering at rate h. The input a_{in} is driven by a coherent probe E_p of frequency ω_p, whereas b_{in} is in a vacuum state. (**B** to **G**) Theoretical results from our model (*24*). (B) to (D) Atomic localization with well-defined azimuthal phase $\varphi = \pi/2$. (B) and (C) Transmission spectra for the steady-state forward flux $T_F(\Delta) = \langle a_{out}^+ a_{out}\rangle / \langle a_{out}^+ a_{out}\rangle_{\Delta>>\kappa}$ and backward flux $T_B(\Delta) = \langle b_{out}^+ b_{out}\rangle / \langle a_{out}^+ a_{out}\rangle_{\Delta>>\kappa}$ as functions of probe detuning $\Delta = \omega_c - \omega_p$. (D) Steady-state intensity correlation function $g_F^{(2)}(\Delta) = \langle (a_{out}^+)^2 (a_{out})^2\rangle / \langle a_{out}^+ a_{out}\rangle^2$ for the forward flux. (E) to (G) Random atomic positions, with an average over $0 \le \varphi < 2\pi$. (E) and (F) Average steady-state transmission spectra $\bar{T}_F(\Delta), \bar{T}_B(\Delta)$. (G) Average steady-state intensity correlation function $\bar{g}_F^{(2)}(\Delta)$. In all frames, the red curves are for no atom ($g_0 = 0$), whereas the blue curves are with one atom coupled as illustrated in (A). The parameters for all plots (B) to (G) are $(g_0, \kappa_i, h, \gamma)/2\pi =$ (70,5,250,1) MHz with $\kappa_{ex} = \kappa_{ex}^{\sigma}$. For $\varphi = 0$, the plots are identical to those in (B) to (D) with the replacement $\Delta \to -\Delta$.

[1]Norman Bridge Laboratory of Physics, 12-33, California Institute of Technology, Pasadena, CA 91125, USA. [2]T. J. Watson Laboratory of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA.

*Present address: Department of Physics, University of Auckland, Auckland, New Zealand.

†Present address: PRESTO, Japan Science and Technology Agency, Saitama, Japan.

‡To whom correspondence should be addressed. E-mail: hjkimble@caltech.edu

tural effect due to anharmonic energy spectra for multiple excitations. The anharmonicity can be exploited to regulate photon transmission [refer to the level diagram in figure 1 of (*14*)] or, alternatively, to spectrally isolate emission events arising from single excitations [as illustrated in the schematic for biexciton decay in figure 1 of (*10*)].

In contrast to transport governed by structure, we describe observations of photon blockade in which photon transport is regulated dynamically by the conditional state of one intracavity atom, leading to an efficient mechanism that is insensitive to many experimental imperfections. As illustrated in Fig. 1, an atom interacting with the fields of a microtoroidal resonator (*20*) regulates the photon statistics of light transmitted and reflected by the resonator. This regulation is achieved by way of an interference effect involving the directly transmitted optical field, the intracavity field in the absence of the atom, and the polarization field radiated by the atom, with the requisite nonlinearity provided by the quantum character of the emission from one atom. Detection of an initial photon in the forward-propagating transmission results in subsequent photons from the incident flux being rerouted. This mechanism requires only that the intracavity atomic absorption be large and thus is robust against variations in atomic coupling. For fiber coupling to and from a microtoroidal resonator (*21*, *22*), we achieved high efficiency for the transport of photons through the turnstile.

Our investigation relied on measurements of photon correlations based on the normalized intensity correlation function $g_F^{(2)}(\tau)$ for the forward-propagating transmitted light (*23*). The incident field (which approximates a coherent state) has $g^{(2)}(\tau) = 1$, corresponding to a Poisson distribution for photon number independent of time delay τ. An ideal photon turnstile would achieve $g_F^{(2)}(0) = 0$ in correspondence to the state of a single photon. More generally, $g_F^{(2)}(0) < 1$ represents a nonclassical effect with the variance in photon number reduced below that of the incident field. For continuous illumination, $g_F^{(2)}(\tau)$ regresses to steady state over a time τ_B, with photon transmissions separated by time intervals $\Delta t >> \tau_B$ being statistically independent. The observation $g_F^{(2)}(0) < g_F^{(2)}(\tau)$ represents photon antibunching; photons are transported one by one through the turnstile without neighbors in time.

The mechanism responsible for photon regulation in our microtoroidal resonator is explained in more detail in Fig. 1. The basic model underlying this analysis is that of a single two-state atom coupled to a resonant cavity, namely the Jaynes-Cummings model extended to incorporate two cavity modes and to include damping by way of reservoir couplings (*24*). As depicted in Fig. 1A, the two internal counterpropagating modes a and b of the toroidal resonator have common frequency ω_C in the absence of scattering. These modes are coupled because of scattering with strength h. The field decay rate for the resonator modes is $\kappa = \kappa_i + \kappa_{ex}$, where κ_i represents intrinsic losses and κ_{ex} describes extrinsic loss due to (adjustable) coupling of the modes to the fiber coupler (*21*, *22*).

We consider an impedance-matched input (that is, critical coupling) for which κ_{ex} is set to $\kappa_{ex}^{cr} = \sqrt{\kappa_i^2 + h^2}$ and the resonator excited with probe detuning $\Delta \equiv \omega_C - \omega_p = 0$. In this case, interference between the cavity field a and the input field a_{in} results in zero flux in the forward direction, $\langle a_{out} \rangle = 0$, where $a_{out} = a_{in} + \sqrt{2\kappa_{ex}}\, a$. The incident flux $|E_p|^2$ appears in the backward direction b_{out} for low internal loss $\kappa_i << \kappa_{ex}$, as shown by the (normalized) probe spectra $T_F(\Delta)$, $T_B(\Delta)$ for the forward- and backward-propagating fields a_{out} and b_{out} in Fig. 1, B and C, with $T_F(\Delta = 0) = 0$ whereas $T_B(\Delta = 0) \approx 1$ (red curves).

An atom near the external surface of the toroid has coherent interactions with the evanescent components of modes a and b that modify T_F and T_B. The fundamental description of these interactions is in terms of normal modes $A = (a + b)/\sqrt{2}$ and $B = (a - b)/\sqrt{2}$, which have mode functions $\psi_{A,B}(\rho, \varphi, z) = f(\rho, z)\{\cos(\varphi), \sin(\varphi)\}$ that are standing waves around the circumference φ of the toroid, with ρ the radial distance from the surface and z the vertical coordinate (*24*). The rate of coherent coupling of an atom to the A and B modes is given by $g_{A,B} = g_0\, \psi_{A,B}(\rho, \varphi, z)$, where for the geometry of our current resonator, we calculate that $g_0/2\pi \approx 95$ MHz (*20*).

Figure 1 shows examples of theoretical probe spectra $T_F(\Delta)$, $T_B(\Delta)$ in the presence of an atom (blue curves) with atom-cavity detuning $\Delta_{AC} \equiv \omega_A - \omega_C = 0$ and for parameters relevant to our current experiment. Relative to the case with $g_0 = 0$ (no atom), there is now a sharp central feature around $\Delta = 0$ associated with atom absorption (*25*), where the width Γ of this feature exceeds the free-space radiative rate γ due to enhanced coupling to the cavity [equation 15 in (*24*)]. For normally ordered

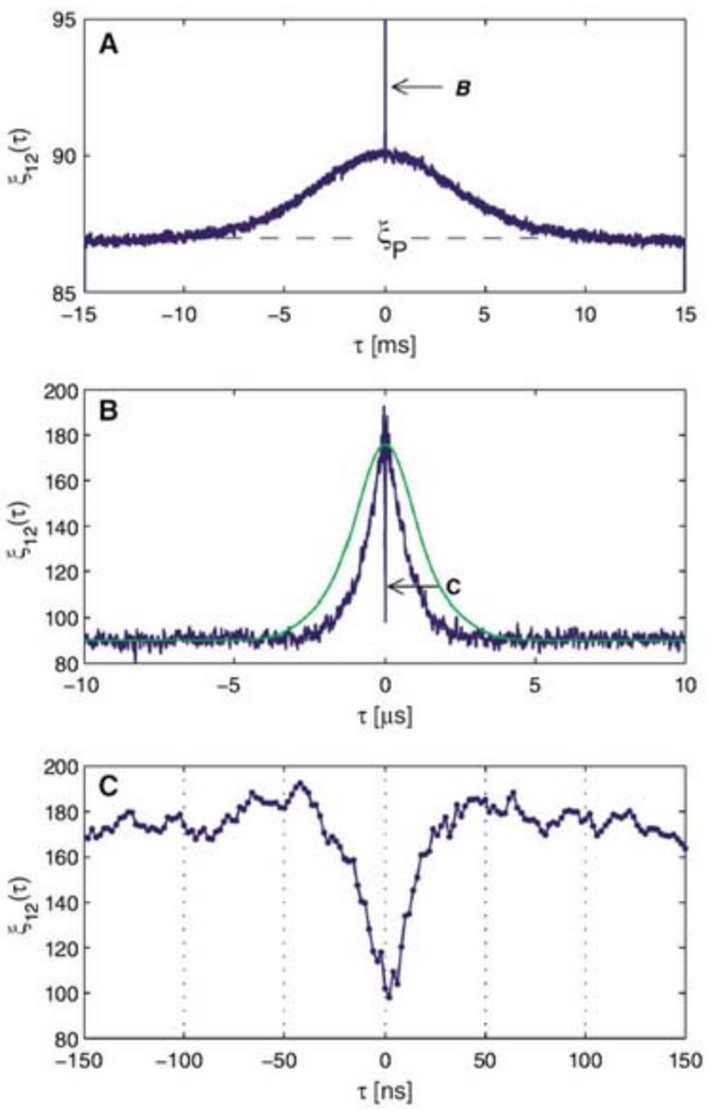


Fig. 2. Cross correlation $\xi_{12}(\tau)$ between photon counts at detectors D_1 and D_2 for the forward flux $\langle a_{out}^{+} a_{out} \rangle$ as a function of time offset τ. (**A**) The broad increase in $\xi_{12}(\tau)$ above the background level ξ_P is the envelope of atom transit events as a cloud of cold Cs atoms falls past the toroidal resonator. (**B**) The center feature indicated by the arrow in (A) is shown on an expanded time scale, revealing the average profile for individual atom transits. For comparison, the green curve is from the cross correlation of $P_1(t)$,$P_2(t+\tau)$ in Fig. 3A. (**C**) The center dip marked by the arrow in (B) is expanded further in τ to reveal a dip in joint detection events due to photon blockade by the atom-cavity system during the atom transit.

expectation values, the following replacements can be made (*24*)

$$a_{out} \rightarrow \alpha_0 + \alpha_-\sigma_-,\ b_{out} \rightarrow \beta_0 + \beta_-\sigma_- \quad (1)$$

where σ_- is the lowering operator for atomic transitions, α_0, β_0 are c numbers derived for the empty cavity with no atom, and $(\alpha_-, \beta_-) \propto g_0$ describes the atom-cavity interaction. For critical coupling, $\alpha_0(\Delta_p \approx 0) \approx 0$, and the steady-state flux $\langle a^+_{out} a_{out}\rangle$ derives principally from the contribution of the atom, given by the term $\alpha_-\sigma_-$ in Eq. 1. In contrast, β_0 is nonzero near critical coupling, with the output b_{out} then determined by the coherent sum of contributions from the terms β_0 and $\beta_-\sigma_-$. For the particular case of atomic interaction with a single normal mode (such as $\varphi = \pi/2$) as in Fig. 1, B to D, these considerations lead to a transmission spectrum for the forward-propagating light a_{out} for which $T_F(\Delta_p \approx 0) \approx 1$, whereas for the backward light b_{out}, $T_B(\Delta_p \approx 0) \approx 0$.

Photon blockade arises in this setting as follows. With critical coupling, $\alpha_0 = 0$, so that a first photon transmitted into a_{out} can originate only from the atom (that is, the term $\alpha_-\sigma_-$ in Eq. 1). This emission projects the atom into its ground state, from which a second emission cannot occur ($\langle \sigma_+^2 \sigma_-^2 \rangle = 0$) (*26*, *27*). A second photon cannot be transmitted until the atomic state regresses to a steady state, which occurs over a time interval set by the cavity-enhanced emission rate Γ.

This qualitative prediction is substantiated in Fig. 1D from a calculation of the intensity correlation function $g_F^{(2)}(\tau = 0)$ for a_{out}. Around $\Delta = \omega_C - \omega_p = 0$, $g_F^{(2)}(0) \approx 0$ as expected for photon blockade. As suggested by Fig. 1C, photons arriving during an initial transmission event will be rerouted to the backward-propagating field b_{out}, so that photon bunching should be expected for b_{out}, as is indeed the case for $g_B^{(2)}(\tau = 0)$ (fig. S3).

Overall, for the conditions considered in Fig. 1, B to D, the atom-cavity system functions as a photon turnstile, with single photons transmitted into a_{out} and excess photons reflected into b_{out}. Conditioned on a first transmission event, the transmission characteristics of the system switch from those in Fig. 1, B and C, for the case $g_0 \neq 0$ to those for $g_0 = 0$. That is, a first photon transmitted triggers the change $\{T_F(\Delta \approx 0) \approx 1, T_B(\Delta \approx 0) \approx 0\}$ for $g_0 \neq 0 \rightarrow \{T_F(\Delta \approx 0) \approx 0, T_B(\Delta \approx 0) \approx 1\}$ with $g_0 = 0$, with regression back to the case $g_0 \neq 0$ occurring on a time scale τ_B.

The case of atomic localization at fixed ϕ as in Fig. 1, B to D, is not realistic for the experiment we performed. Hence, in Fig. 1, E to G, we present corresponding results averaged over the azimuthal angle φ (that is, for single-atom interactions but each at a random location in φ). The dominant effect is a reduction in the contrast for transmission and reflection. However, the atom-cavity system still functions effectively as a photon turnstile for the forward flux transmitted into a_{out}, with $\bar{g}_F^{(2)}(0) \approx 0$ at $\Delta = 0$, where $\bar{g}_F^{(2)}$ denotes an azimuthal average. For further discussion, see (*24*).

Our experiment to investigate these effects was similar to the setup shown in figure 1 of (*20*), with further details given in section I of (*24*). The principal difference between the depiction in Fig. 1A and the actual experiment is that individual Cs atoms were not trapped in the evanescent field of the toroid, but rather fell along z and transited through the evanescent field of the resonator. The toroidal resonator was monolithically fabricated from SiO_2 on a Si chip (*28*), had major diameter $D \approx 25$ μm and minor diameter $d \approx 6$ μm, and was located within a chamber at ultrahigh vacuum. From measurements of atom transit events as functions of atom-cavity detuning Δ_{AC} and intracavity photon number $\langle n \rangle$, we determined a single effective coupling strength $g_{eff}/2\pi \approx 70$ MHz (*24*), where ω_A corresponds to the $6S_{1/2}, F = 4 \rightarrow 6P_{3/2}, F' = 5$ transition in atomic Cs.

In contrast to our report in (*20*), we were not in the strong coupling limit of cQED, specified by $g_{eff} >> (\kappa^{cr}, \gamma_\perp)$, because the rate of decay for the cavity field was $\kappa^{cr}/2\pi = (165 \pm 15)$ MHz, whereas the free-space rate of decay of the atomic polarization was $\gamma_\perp/2\pi = 2.6$ MHz. Rather, the conditions were more appropriate to the "bad-cavity"

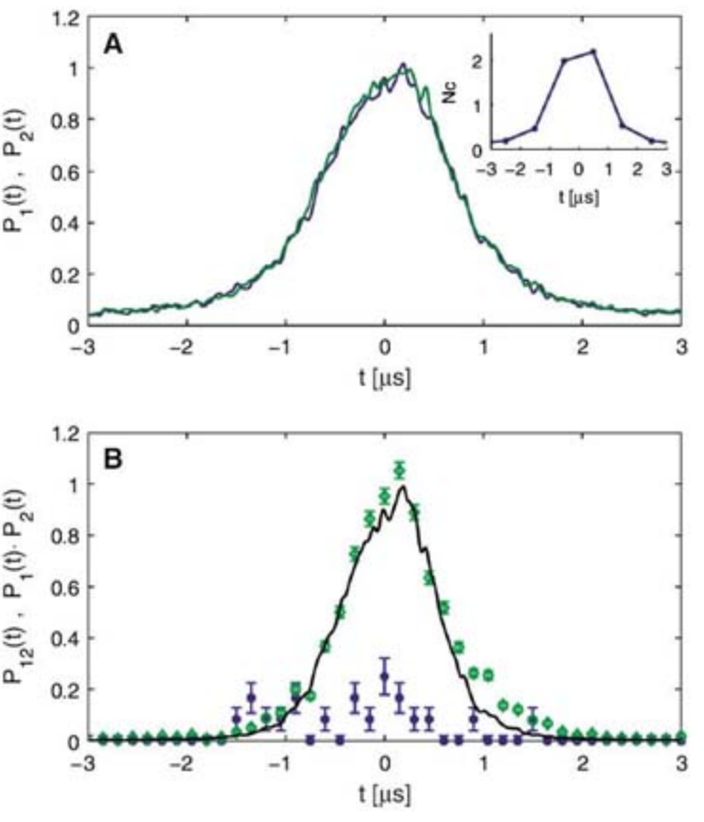


Fig. 3. Single $P_{12}(t)$ and joint $P_{12}(t)$ event probabilities versus time t around the center of transit events. (**A**) $P_1(t)/P_1(0)$ and $P_2(t)/P_2(0)$, where $P_1(0) = 2.88 \times 10^{-3}$, $P_2(0) = 2.80 \times 10^{-3}$. The inset shows the average number of counts N_c recorded during a transit for time bins $\delta t = 0.5$ μs. (**B**) $P_{12}(t)/[P_1(0)P_2(0)]$ (blue circles) and $[P_1(t)P_2(t)]/[P_1(0)P_2(0)]$ (black trace). There is a clear nonclassical suppression $P_{12}(t) < P_1(t)P_2(t)$. The green diamonds are the average of $P_{12}(t, t \pm \tau) = \langle p_1(t) p_2(t + \tau)\rangle / \langle p_1(t)\rangle \langle p_2(t + \tau)\rangle$ for 26 ns $\leq \tau \leq$ 44 ns and demonstrate that emission events become statistically independent for $\tau > \tau_B$.

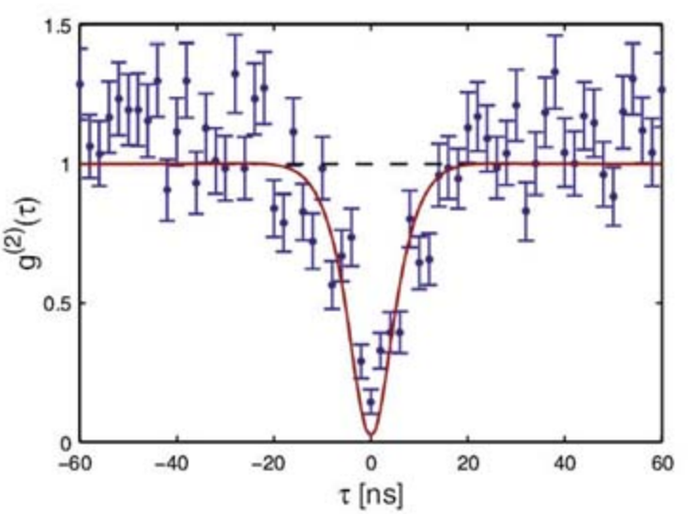


Fig. 4. Intensity correlation function $g^{(2)}(\tau)$ versus time delay τ for individual atomic transit events. $g^{(2)}(\tau)$ exhibits photon antibunching $g^{(2)}(0) < g^{(2)}(\tau)$, and sub-Poissonian photon statistics $g^{(2)}(0) = (0.14 \pm 0.04) < 1$, over an interval $\Delta\tau \approx 6$ ns (half width at half maximum) due to the operation of the atom-cavity system as a photon turnstile. The red trace is from our theoretical model.

limit, for which $\kappa^{cr} >> (g_{eff})^2/\kappa^{cr} >> \gamma_\perp$ (*25*). Other measured parameters for our system were $(\kappa_i, \kappa_{ex})/2\pi = (75, 90)$ MHz for the case of critical coupling.

For critical coupling with $\Delta \approx 0 \approx \Delta_{AC}$, we directed the forward flux for a_{out} (Fig. 1A) to two single-photon counters $D_{1,2}$ (fig. S1), with the outputs $C_{1,2}(t_i)$ from these detectors being time-stamped and stored for time bins $\delta t = 2$ ns for each value t_i. From raw records for $C_{1,2}(t_i)$, we directly computed the cross correlation $\xi_{12}(\tau) = \sum_i \langle C_1(t_i) C_2(t_i + \tau) \rangle$, with an example given in Fig. 2. Clearly evident in Fig. 2A is a broad bump of width $\Delta\tau_{cloud} \approx 7.5$ ms, corresponding to the overall envelope for transit events as the cloud of cold atoms falls past the toroid. The observed value for $\Delta\tau_{cloud}$ is in good agreement with the value of ~7 ms expected for our geometry (*24*). The pedestal at level ξ_P in Fig. 2A arose from background counts and was unchanged in the absence of atoms.

The central 20-μs interval in Fig. 2A is expanded in Fig. 2B to display a peak of full width $\Delta\tau_{transit} \approx 2$ μs that arises from transit events for individual atoms falling through the evanescent field of the toroid. Each atom experiences a time-varying coupling $g_0(\rho, \varphi, vt)$, where v is the atomic velocity in the z direction, leading to a temporal variation in the forward flux $T_F \propto \langle a^+_{out} a_{out} \rangle$. Again, the observed time dependence is in accord with a theoretical model based on the calculated mode function for our toroid and the known atomic velocity v [see figure 3D in (*20*)].

Finally, the central 300-ns interval in Fig. 2B is expanded in Fig. 2C to reveal a dip in $\xi_{12}(\tau)$ around $\tau = 0$, corresponding to a paucity of coincidence counts between detectors D_1 and D_2 over an interval $\Delta\tau \approx 22$ ns (full width at half maximum). This suppressed probability for joint detection agrees with that expected from the analysis in Fig. 1 for a photon turnstile. However, independent of any purported microscopic mechanism, the results presented in Fig. 3C for $\xi_{12}(\tau)$ represent a prima facie observation of nonclassical light from the toroid, because $\xi_{12}(0) < \xi_{12}(\tau)$ (*23*). However, these results do not exhibit sub-Poissonian photon statistics. $\xi_{12}(0)$ does not drop below the level of the pedestal ξ_P shown in Fig. 2, A and B, because of the random arrival of atoms into the cavity mode (*29*).

Figure 2 is from an unconditional analysis of the records of photo counts $C_{1,2}(t_i)$. Because we could identify transit events for single atoms with high statistical confidence (*20*, *24*), we next performed an analysis conditioned on the presence of an atom. As described in section II of (*24*), for individual atomic transits recorded by detectors $D_{1,2}$, we derive the probabilities $p_{1,2}(t_i)$ and $p_{12}(t_i)$ for single and joint detections at $D_{1,2}$ during an atomic transit.

Figure 3A displays the single-event probabilities $P_{1,2}(t) \equiv \langle p_{1,2}(t) \rangle$ as functions of time t around the center $t = 0$ of the transit events with time bins of duration $\delta t = 2$ ns. The time dependence for $P_{1,2}(t)$ agrees with an independent estimate based on our cavity mode and the atomic velocity. The assertion that the peak in Fig. 3A arises predominantly from actual transit events is further supported by a comparison with the central feature in Fig. 2B.

Figure 3B presents results for the photon statistics for these transit events, now with $\delta t = 150$ ns. The joint probability $P_{12}(t) \equiv \langle p_1(t) p_2(t) \rangle$ lies well below the level set by $P_1(t)P_2(t)$, violating the Cauchy-Schwarz inequality with $P_{12} < P_1 P_2$. Over the central 1-μs interval, we find that $P_{12}/P_1P_2 = (0.135 \pm 0.033) < 1$. In agreement with our theoretical analysis, an atomic transit event regulates the photon statistics of the output field, with a_{out} becoming markedly sub-Poissonian because of the atom-cavity interaction in the fashion of a photon turnstile.

Final support for this interpretation comes from Fig. 4, where we present measurements of the intensity correlation function $g^{(2)}(t, \tau) \equiv \langle p_1(t) p_2(t + \tau) \rangle / \langle p_1(t) \rangle \langle p_2(t + \tau) \rangle$. For times t_0 such that $-0.4\ \mu s \leq t_0 \leq +0.4\ \mu s$ around the center of the transit events in Fig. 3, $g^{(2)}(\tau) \equiv g^{(2)}(t_0, \tau)$ exhibits clear antibunching, as well as sub-Poissonian photon statistics. The full trace in Fig. 4 is from the steady-state solution to the master equation from our theoretical model, which has been averaged over the azimuthal angle φ [section VIII of (*24*)]. All parameters for this comparison were determined from independent measurements, including the effective coupling strength g_{eff}, which was used for the theoretical trace in Fig. 4. The agreement between theory and experiment is evidently quite reasonable. In particular, photon blockade persists over a time τ_B set by the cavity-enhanced atomic decay rate Γ, which is approximately given by $\tau_B \approx 2.5/\Gamma \approx 7$ ns for the parameters of our experiment [eq. 16 of (*24*)], in good agreement with $\Delta\tau \approx 6$ ns in Fig. 4.

Our results represent an observation of manifestly quantum (nonclassical) fields for cQED with single atoms and microresonators. The underlying dynamical mechanism for photon blockade is quite robust against many experimental imperfections, including variations in the coupling $g(\vec{r})$, requiring only that $g^2/\kappa^{cr}\gamma_\perp > 1$. For comparison, in our experiment $(g_{eff})^2/\kappa^{cr}\gamma_\perp \approx 11$. This mechanism is also robust against stronger excitation; numerical solutions of the full master equation show that the turnstile effect persists for an intracavity photon number up to the order of unity in the absence of an atom. Moreover, the measured throughput efficiency ζ for single photons [$T_F(\Delta = 0)$ in Fig. 1, B and E] reaches $\zeta \approx 25\%$ during single-atom transit events. Our theoretical analysis suggests that it should be possible to achieve $\zeta > 90\%$, thereby realizing an efficient photon turnstile for which an input field is sorted by photon number into forward- and backward-propagating output fields with small loss. By operating in a pulsed domain, investigations of photon collisions become possible, as well as quantum nondemolition detection. Because our toroidal resonators are lithographically fabricated with input-output coupling via optical fiber (*22*), our experiment provides an important first step into the quantum domain to implement scalable atom-cavity systems, including for quantum logic with photons (*30*) and for quantum processes on atom chips (*31*).

References and Notes

1. T. A. Fulton, G. J. Dolan, *Phys. Rev. Lett.* **59**, 109 (1987).
2. M. A. Kastner, *Rev. Mod. Phys.* **64**, 849 (1992).
3. K. K. Likharev, *Proc. IEEE* **87**, 606 (1999).
4. P. Zoller *et al.*, *Eur. Phys. J. D* **36**, 203 (2005).
5. R. Miller *et al.*, *J. Phys. B At. Mol. Opt. Phys.* **38**, S551 (2005).
6. T. Wilk, S. C. Webster, A. Kuhn, G. Rempe, *Science* **317**, 488 (2007).
7. H. Walther, *Fortschr. Phys.* **52**, 1154 (2004).
8. J. M. Raimond *et al.*, *J. Phys. B At. Mol. Opt. Phys.* **38**, S535 (2005).
9. For a review, see (*32*).
10. For a review, see (*33*).
11. J. Majer *et al.*, *Nature* **449**, 443 (2007).
12. M. A. Sillanpää, J. I. Park, R. W. Simmonds, *Nature* **449**, 438 (2007).
13. A. Imamoglu *et al.*, *Phys. Rev. Lett.* **79**, 1467 (1997).
14. K. M. Birnbaum *et al.*, *Nature* **436**, 87 (2005).
15. L. Tian, H. J. Carmichael, *Phys. Rev. A* **46**, R6801 (1992).
16. J. Kim *et al.*, *Nature* **397**, 500 (1999).
17. P. Grangier, D. F. Walls, K. M. Gheri, *Phys. Rev. Lett.* **81**, 2833 (1998).
18. M. J. Werner, A. Imamoglu, *Phys. Rev. A* **61**, 011801 (1999).
19. S. Rebić, A. S. Parkins, S. M. Tan, *Phys. Rev. A* **65**, 043806 and 063804 (2002).
20. T. Aoki *et al.*, *Nature* **443**, 671 (2006).
21. M. L. Gorodetsky, A. D. Pryamikov, V. S. Ilchenko, *J. Opt. Soc. Am. B* **17**, 1051 (2000).
22. S. M. Spillane, T. J. Kippenberg, O. J. Painter, K. J. Vahala, *Phys. Rev. Lett.* **91**, 043902 (2003).
23. L. Mandel, E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, 1995).
24. Materials and methods are available as supporting material on *Science* Online.
25. Q. A. Turchette, R. J. Thompson, H. J. Kimble, *Appl. Phys. B* **60**, S1 (1995).
26. H. J. Kimble, M. Dagenais, L. Mandel, *Phys. Rev. Lett.* **39**, 691 (1977).
27. H. J. Carmichael, *Phys. Rev. Lett.* **55**, 2790 (1985).
28. D. K. Armani *et al.*, *Nature* **421**, 925 (2003).
29. H. J. Kimble, *Phys. Rev. Lett.* **90**, 249801 (2003).
30. L.-M. Duan, H. J. Kimble, *Phys. Rev. Lett.* **92**, 127902 (2004).
31. P. Treutlein *et al.*, *Fortschr. Phys. Progr. Phys.* **54**, 702 (2006).
32. G. Khitrova *et al.*, *Nat. Phys.* **2**, 81 (2006).
33. A. J. Shields, *Nat. Photon.* **1**, 215 (2007).

34. We gratefully acknowledge the contributions of S. Kelber, J. Petta, C. Regal, and E. Wilcut-Connolly. This research is supported by NSF, the Intelligence Advanced Research Projects Activity, and Northrop Grumman Space Technology.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1062/DC1
Materials and Methods
Figs. S1 to S4
References

29 October 2007; accepted 9 January 2008
10.1126/science.1152261

The Force Needed to Move an Atom on a Surface

Markus Ternes,[1]† Christopher P. Lutz,[1] Cyrus F. Hirjibehedin,[1]*
Franz J. Giessibl,[2] Andreas J. Heinrich[1]

Manipulation of individual atoms and molecules by scanning probe microscopy offers the ability of controlled assembly at the single-atom scale. However, the driving forces behind atomic manipulation have not yet been measured. We used an atomic force microscope to measure the vertical and lateral forces exerted on individual adsorbed atoms or molecules by the probe tip. We found that the force that it takes to move an atom depends strongly on the adsorbate and the surface. Our results indicate that for moving metal atoms on metal surfaces, the lateral force component plays the dominant role. Furthermore, measuring spatial maps of the forces during manipulation yielded the full potential energy landscape of the tip-sample interaction.

In 1990, Eigler and Schweizer (*1*) positioned single Xe atoms with atomic-scale precision on a Ni(110) surface using a scanning tunneling microscope (STM). This technique of atom manipulation has subsequently been used to build model physical systems—such as quantum-confined structures (*2*, *3*), magnetic nanostructures (*4*, *5*), and artificial molecules (*6*, *7*)—one atom at a time. In the most common STM manipulation technique, the adsorbate follows the tip by hopping from one surface binding site to the next, because a partial chemical bond is formed between the apex atom of the tip and the adsorbed atom or molecule. Previous studies of the manipulation process with STM (*8*, *9*) were highly revealing but could not determine the forces involved in manipulation. Recently, the atomic force microscope (AFM) has been used to manipulate atoms at semiconducting surfaces (*10*), even at room temperature (*11*, *12*). Atom manipulation with an AFM is particularly promising, because it allows the direct measurement of the required forces.

In this work, we used an AFM to quantify the forces required to pull individual adsorbates along a surface. We find that moving cobalt (Co) on Pt(111) requires a lateral force of 210 pN and that this force is independent of the vertical force. The lateral force can vary substantially with the chemical nature of the underlying surface as it is only 17 pN for Co on Cu(111). For both surfaces, the force on the tip due to the Co atom is nearly spherically symmetric. In contrast, for manipulating a carbon monoxide (CO) molecule, the forces are more complex, deviating markedly from spherical symmetry.

We used a frequency modulated AFM (*13*) with the qPlus sensor design (*14*) operating in ultrahigh vacuum at a temperature of ~5 K. A metal tip was mounted on an oscillating cantilever (resonance frequency f_0 = 21,860 Hz) and used to probe the surface and move the adsorbates (Fig. 1A). The stiff cantilever (spring constant $k_0 \approx 1800$ Nm^{-1}) allows stable, small amplitude oscillation (A = 30 pm) close to the surface. We monitored the shift of the oscillation frequency Δf, which for small A is roughly proportional to the vertical stiffness $k_z \approx 2k_0/f_0 \times \Delta f$ of the tip-sample junction (*13*). To allow comparisons with STM manipulation experiments, we also detected the tunneling conductance G between tip and surface (*15*).

Figure 1, B to E, shows constant-height images of a single Co atom on Pt(111) and a single CO molecule on Cu(111) with tip heights close to the threshold for atom manipulation. In these images, k_z and G show circular symmetry, without any sign of the threefold substrate symmetry. The images obtained from Co on Pt(111) show a narrow dip in k_z and a peak in G. In contrast, the images from a CO molecule on Cu(111) have a more complex structure: k_z is flat around the central minimum, and G contains a central conductance peak within a broader conductance dip. For an asymmetric tip apex, k_z often deviates from circular symmetry for CO (*16*). In these images, the change in stiffness due to the adsorbate is 17 Nm^{-1} on Co and 9 Nm^{-1} on CO, in the range of a metal-metal bond stiffness (10 to 100 Nm^{-1}).

We can derive the force to move an atom from the measurement of k_z as a function of both vertical and lateral tip position. Figure 2 shows "line scans" obtained by moving the tip parallel to the surface at constant height, passing over the top of an isolated Co atom on Pt(111). These scans were repeated at progressively smaller tip heights until the Co atom hopped to a neighboring adsorption site, as illustrated in Fig. 2A. The direction of these scans corresponds to the direction of easiest motion on this substrate: It connects two neighboring threefold hollow adsorption sites. Figure 2, C and D, shows that both G and $|k_z|$ increased as the tip height was decreased. At the smallest tip height and at a lateral position about halfway to the adjacent binding site, a sudden jump occurred in both k_z and G. At this tip position, the Co atom reproducibly hopped from its initial binding site to the next and was imaged again at its new position by the continuing line scan. In contrast, the atom reliably remained at its initial binding site when the tip was only 5 pm farther from the surface.

The vertical force F_z was determined by integrating k_z along z. The frequency shift, and therefore k_z, is an average of the tip-surface interaction over one oscillation cycle of the tip. This vertical blurring in all data was removed when computing F_z from k_z, by means of a deconvolution process (*17*).

We interpret F_z as the sum of two components: a background force F_B and the force F_z^* due to the presence of the adsorbate (Fig. 2B).

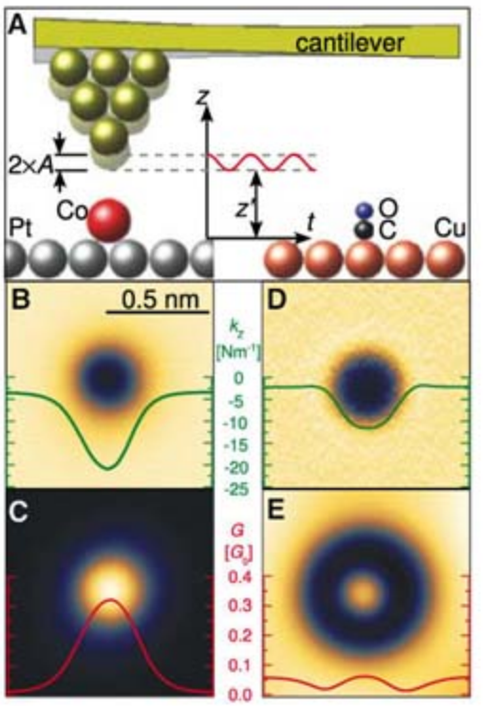


Fig. 1. Simultaneous AFM and STM measurements of individual adsorbates. (**A**) An atomically sharp metal tip is oscillating in z with an amplitude A = 30 pm over a flat metal surface on which an individual Co atom or CO molecule is adsorbed. The measured frequency shift of the cantilever from its natural resonance frequency is proportional to the vertical stiffness k_z of the tip-sample interaction. A small bias voltage of 1 mV was applied between tip and sample to measure the tunneling current, which is proportional to the conductance G, given in units of the single-channel, spin-degenerate quantum of conductance $G_0 = 2e^2/h$ = (12,906 ohm)$^{-1}$, where e is the elementary charge and h is Planck's constant. The inset graph shows the tip motion $z(t)$ between its closest distance (z') and farthest distance ($z' + 2A$) from the sample. The ball models of the surfaces are scaled to match the dimensions of the images in the following panels. (B to E) Images measured at a constant height z' close to the threshold for adsorbate motion. (**B**) The tip-sample stiffness of a single Co atom on Pt(111). (**C**) Tip-sample conductance measured simultaneously for the same system as in (B). (**D** and **E**) Same as (B) and (C) for a single CO molecule on Cu(111). The colored curves in the panels are horizontal cross sections through the centers of the images.

[1]IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA. [2]Institute of Experimental and Applied Physics, University of Regensburg, D-93053 Regensburg, Germany.

*Present address: London Centre for Nanotechnology, Department of Physics and Astronomy, Department of Chemistry, University College London, London WC1H 0AH, UK.

†To whom correspondence should be addressed. E-mail: markust@us.ibm.com

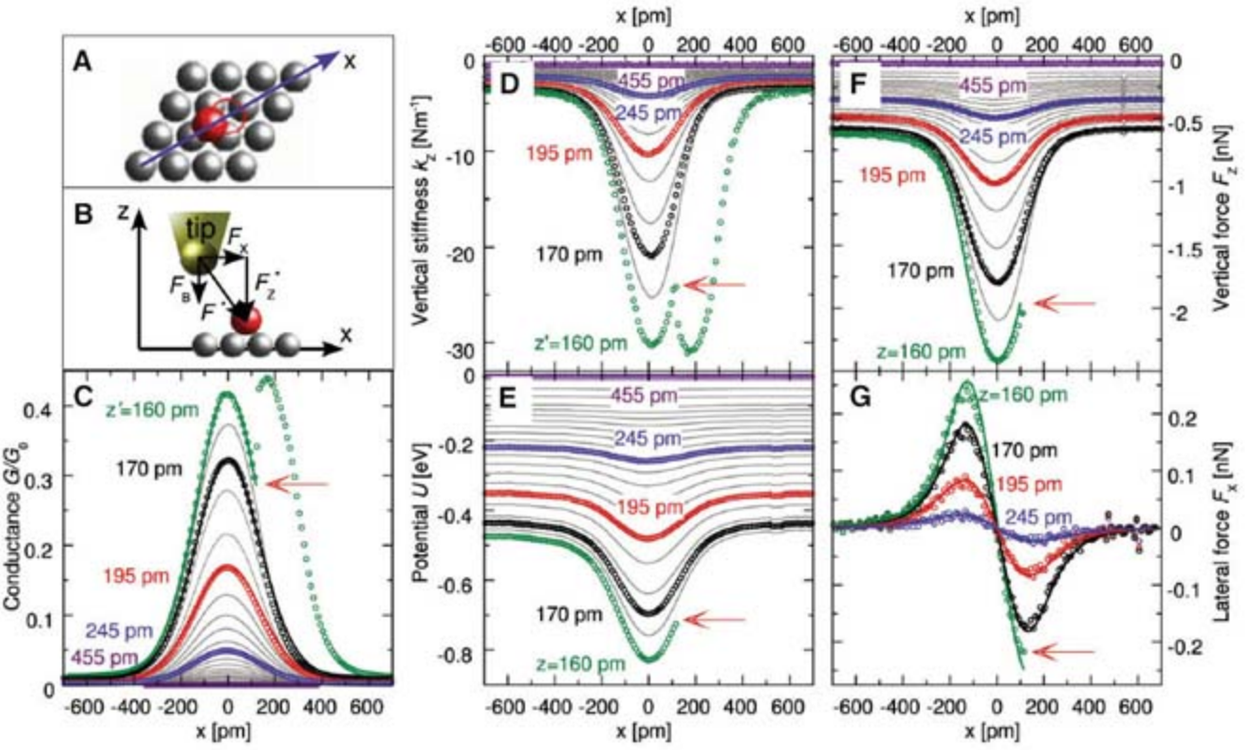


Fig. 2. Measuring the force to move Co on Pt(111). (**A**) Schematic top view of the Pt(111) surface atoms (gray) and the adsorbed Co atom (red). In the following panels, constant-height line scans in the direction of easiest adsorbate motion (x direction) were taken at successively reduced tip-sample separations until the Co atom hopped to the adjacent adsorption site [red circle in (A)]. The scan speed was ~0.5 nm/s. (**B**) The force F^* between tip apex and the Co atom can be divided into the lateral force F_x and the vertical force F_z^*. The total vertical force F_z is the sum of F_z^* and the background force F_B. (**C** and **D**) Simultaneously measured conductance G and stiffness k_z (circles and gray lines). Note that these values are time-averaged over the cantilever oscillation between $z = z'$ and $z = z' + 2A$. We label selected line scans with the closest approach z' during the oscillation (*15*). (**E** to **G**) Tip-sample interaction energy U, vertical force F_z, and lateral force F_x extracted from the stiffness k_z data in (D). Selected line scans are labeled with the tip height z; here, the oscillation amplitude has been deconvolved from the curves. The red arrows in (C) to (G) indicate the hop of the Co atom to the neighboring binding site. Colored lines in (C), (F), and (G) are fits with the s-wave model.

The background force is in large part due to long-range (van der Waals) forces (*18*) and increases as the tip approaches the surface but does not depend on the lateral position. The vertical force F_z^* due to the adsorbate grows rapidly (Fig. 2F and fig. S3B), doubling in magnitude as z is changed by only 15 pm near the move threshold, indicating the short-range nature of this force.

We found the tip-sample interaction force to be nondissipative as long as the adsorbate did not hop to a new binding site (*19*). In this nondissipative range of tip positions, we calculated the tip-sample interaction potential U (Fig. 2E) by integration of F_z along z (*20*). The lateral force F_x was then calculated by differentiation of U in the x direction (Fig. 2G). This technique allowed us to determine forces in any direction, even though the cantilever only senses the vertical stiffness. The lateral force was zero with the tip placed above the adsorbate, grew as the tip was moved laterally until a maximum was reached, and vanished far from the adsorbate.

This procedure allows us to determine the force that it takes to move a single Co atom across a Pt(111) surface. At the point where the atom hopped, we found a lateral force $F_x = 210 \pm 30$ pN (*21*). This force varied by only ±5% between different tips (*22*). At the same time, $F_z^* = -1.4 \pm 0.2$ nN was much larger than the lateral force and was nearly half as large as the bond-breaking force of 4 nN for a Pt point contact (*23*). To understand the interplay between the vertical and horizontal force components, we decreased z below the threshold height for hopping and continued to measure the forces. We found that the lateral force to move the atom remained constant, whereas the vertical force varied by a large factor (figs. S1 and S2). For the range of heights measured, F_z^* at the point where the Co atom hopped varied from −0.45 nN (with the tip laterally far from the atom) to −3.0 nN (with the tip nearly above the atom). These results suggest that the lateral force is the key for the manipulation of metal adsorbates on flat metal surfaces. This insensitivity to F_z^* is in contrast to the mechanism determined for moving Si atoms on Si(111) (*12*). There, it was found that the vertical force plays a dominant role by causing a reduction of the energy barrier between two adsorption sites as a result of relaxation of the Si adsorbate and surface.

The force that is required to move an atom strongly depends on the supporting substrate.

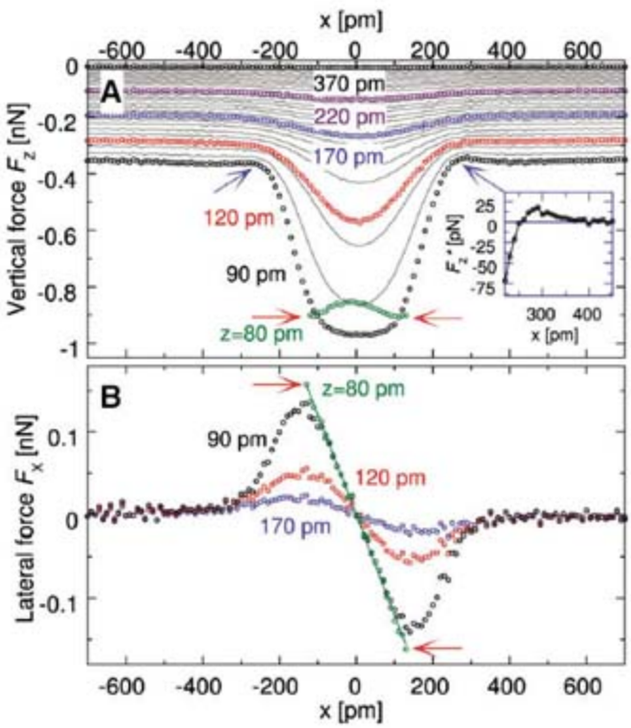


Fig. 3. Vertical and lateral forces for manipulating CO on Cu(111). Vertical (**A**) and lateral (**B**) force components when moving the tip over a CO molecule on Cu(111) for different tip heights z. The x direction corresponds to the easiest adsorbate motion from an on-top binding site to an adjacent one via a bridge site. The blue arrows mark a repulsive force $F_z^* = F_z - F_B$ of up to 20 pN between tip and molecule (see inset). At z = 80 pm, the molecule hopped between neighboring binding sites (red arrows). The green line in (B) is a linear fit to the lateral force with a slope of 1.2 N/m.

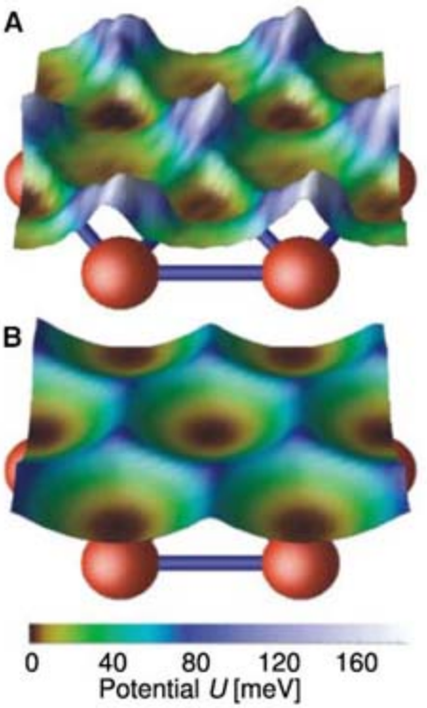


Fig. 4. Tip-adsorbate energy landscape during manipulation. Two-dimensional potential landscapes of the tip-sample interaction energies during controlled manipulation of Co (**A**) and CO (**B**) on Cu(111). The energy scales of the color-coded images are shifted so that $U = 0$ at the preferred adsorption site for Co (fcc hollow site) and CO (on-top site). The underlying Cu(111) lattice is superimposed as a ball-and-stick model. The size of each image is 550 × 480 pm^2. (For more details, see fig. S5.)

Much smaller forces were sufficient to manipulate Co atoms on Cu(111) (fig. S3). Here, the required lateral force was only 17 ± 3 pN, even though Cu and Pt are both face-centered cubic (fcc) crystals and the Co atom binds at a threefold hollow site on both surfaces. This indicates that the nature of the chemical bonding plays a strong role. For Cu, the bonding is dominated by hybridization of the electronic states of the Co adsorbate with the 4*s* metal band, which shows no discernible direction dependence. In contrast, extra bonding occurs on Pt resulting from its partially filled and strongly directional 5*d* bonds (*24*), which apparently increase the forces necessary for manipulation.

To explore the spatial symmetry of the tip-adsorbate force, we modeled it as depending only on the tip-adsorbate distance [for details, see the supporting online material (SOM) text]. For this model, which is similar to the one suggested by Braun and Hla (*25*), the distance dependence of the force was derived from F_z^* measured at $x = 0$ (fig. S4). Despite the simplicity of this model, it agrees well with Co on Pt(111), as shown in Fig. 2, and with Co on Cu(111), as shown in fig. S3. This observed spherical symmetry of the force between the Co adsorbate and the tip apex suggests that the interaction occurs primarily via *s*-wave orbitals in both the tip-apex atom and the Co atom.

The manipulation forces for Co and a small molecule (CO) differed dramatically even though both adsorbates move on Cu(111) at a similar tunneling conductance in STM experiments (*8*). We found that the lateral manipulation force for CO molecules (160 ± 30 pN) is an order of magnitude larger than that for Co atoms (Fig. 3 and fig. S3). More importantly, the spatial dependence of the forces was markedly different. For example, F_z^* at closest tip-sample approach before hopping was almost independent of the lateral tip position around the center of the molecule and became repulsive at $x \approx \pm 300$ pm (Fig. 3). This dependence is in contrast to the *s*-wave nature of the forces that we found for metal adsorbates.

At small tip heights, the adsorbate follows the tip from binding site to binding site (*8*, *9*). Under these conditions, maps of the forces and the tip-sample interaction potential can be constructed by combining k_z images obtained at various tip heights both above and below the manipulation threshold (*26*). Figure 4A shows this interaction potential for Co manipulated on Cu(111). The most stable adsorption sites are the fcc hollow sites (*9*), which appear as minima in the potential map. The neighboring hexagonal close-packed (hcp) sites have a slightly higher potential energy. We note that the Co atom can only be stabilized on this binding site when the tip is in close proximity (*9*). The fcc and hcp sites were separated by a potential barrier of 35 ± 5 meV, whereas the highest potential of 160 ± 30 meV occurred when the tip was placed above an on-top site of the surface. In contrast, in the potential landscape of Co on Pt(111), the barrier is 200 ± 30 meV and the two types of threefold hollow sites are essentially indistinguishable (fig. S5E). The potential landscape for an adsorbed CO molecule (Fig. 4B) clearly reflects the different symmetry of the binding site. In this case, we found that the bridge site had the lowest barrier (70 ± 10 meV) between two on-top binding sites.

For all three systems, the measured potential barrier height for positioning the tip between two neighboring adsorption sites is in close agreement with the diffusion barrier for adsorbate motion as determined by density functional theory [37 meV for Co on Cu(111)] (*27*) and experiments performed without the presence of a probe tip [75 meV for CO on Cu(111) (*28*) and 200 ± 10 meV for Co on Pt(111) (*29*)]. Although diffusion experiments detect the lowest barrier of the adsorbate-surface potential landscape without the presence of a tip, our measurements determine the potential for moving the tip resulting from all interactions among tip, surface, and adsorbate.

The present method for measuring the full tip-sample potential landscape and the forces necessary to manipulate atoms and molecules in arbitrary directions provides important information about the manipulation process without relying on advanced simulations such as density functional theory. It could give additional impetus to the exploration of atomic-scale friction and atom and molecule diffusion on surfaces and offer a deeper insight into controlled bottom-up assembly mechanisms. A systematic investigation of the manipulation forces on different surface-adsorbate combinations is now possible, and the driving mechanism to create future nanoscale devices can be explored in a quantitative manner.

References and Notes

1. D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524 (1990).
2. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Science* **262**, 218 (1993).
3. N. Nilius, T. M. Wallis, W. Ho, *Science* **297**, 1853 (2002).
4. W. Chen, T. Jamneala, V. Madhavan, M. F. Crommie, *Phys. Rev. B* **60**, R8529 (1999).
5. C. F. Hirjibehedin, C. P. Lutz, A. J. Heinrich, *Science* **312**, 1021 (2006).
6. S.-W. Hla, L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **85**, 2777 (2000).
7. R. Yamachika, M. Grobis, A. Wachowiak, M. F. Crommie, *Science* **304**, 281 (2004).
8. L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **79**, 697 (1997).
9. J. A. Stroscio, R. J. Celotta, *Science* **306**, 242 (2004).
10. N. Oyabu, Y. Sugimoto, M. Abe, Ó. Custance, S. Morita, *Nanotechnology* **16**, S112 (2005).
11. Y. Sugimoto *et al.*, *Nat. Mater.* **4**, 156 (2005).
12. Y. Sugimoto *et al.*, *Phys. Rev. Lett.* **98**, 106104 (2007).
13. T. R. Albrecht, P. Grütter, D. Horne, D. Rugar, *J. Appl. Phys.* **69**, 668 (1991).
14. F. J. Giessibl, *Appl. Phys. Lett.* **76**, 1470 (2000).
15. Detecting G allows us to assign $z = 0$ throughout this work as the extrapolated tip height, where the conductance to the bare surface is equal to G_0 (fig. S3A). This approach provides a convenient reference for z, which corresponds roughly to the point contact between the tip and the bare surface. This assignment does not necessarily provide direct comparison of the tip heights between the two surfaces explored in this work. The measured conductance is an average over the tip oscillation. For the amplitude $A = 30$ pm used here, the conductance differs by only ~11% from the value of a nonoscillating tip at the same mean height.
16. F. J. Giessibl, S. Hembacher, H. Bielefeld, J. Mannhart, *Science* **289**, 422 (2000).
17. J. E. Sader, S. P. Jarvis, *Appl. Phys. Lett.* **84**, 1801 (2004).
18. S. Morita, R. Wiesendanger, E. Meyer, Eds. *Noncontact Atomic Force Microscopy* (Springer, Berlin, 2002).
19. We corroborate this assumption with measurements of the power dissipation. Dissipative components in the tip-sample interaction cause an increase in the power required to maintain a constant oscillation amplitude. In our experiment, the energy stored in the mechanical motion of the tip is only $E = \frac{1}{2}k_0A^2 \approx 5$ eV. The cantilever has a mechanical quality factor of $Q \approx 35{,}000$, so that the intrinsic energy loss per oscillation cycle [$= 2\pi E/Q$, see (*18*)] is less than 1 meV. Therefore, we estimate that the dissipative interaction between tip and sample is less than 1% of the typical conservative interaction, so it can be safely neglected; see fig. S5 for more information.
20. H. Hölscher, S. M. Langkat, A. Schwarz, R. Wiesendanger, *Appl. Phys. Lett.* **81**, 4428 (2002).
21. The ±10% uncertainty of the cantilever stiffness produces a systematic error. In addition, the uncertainties given in the text include the error due to the finite tip-height increments (±10%) and the effect that different tips (*22*) produce slightly different results (±5%).
22. Throughout this work, we refer to a changed atomic arrangement at the tip apex (e.g., by touching the tip to the surface) as a different tip.
23. S. R. Bahn, K. W. Jacobsen, *Phys. Rev. Lett.* **87**, 266101 (2001).
24. B. Hammer, J. K. Nørskov, *Nature* **376**, 238 (1995).
25. K.-F. Braun, S.-W. Hla, *Phys. Rev. B* **75**, 033406 (2007).
26. For each lateral tip position, the k_z images for heights above the manipulation threshold were laterally shifted to the location of the binding site nearest the tip. Because dissipation occurred when the atom hopped, comparison of energies between different binding sites might not be valid, but energies within the basin around any one binding site are correct. For more details, see fig. S5.

27. D. V. Tsivlin, V. S. Stepanyuk, W. Hergert, J. Kirschner, *Phys. Rev. B* **68**, 205411 (2003).
28. K. L. Wong, B. V. Rao, G. Pawin, E. Ulin-Avila, L. Bartels, *J. Chem. Phys.* **123**, 201102 (2005).
29. P. Buluschek, thesis, École Polytechnique Fédéral de Lausanne, Switzerland (2007).

30. We thank B. J. Melior, G. Zeltzer, and M. Breitschaft for expert technical assistance and A. F. Otte, D. M. Eigler, A. Schwarz, J. Mannhart, and W. Chaisangmongkon for stimulating discussions. We acknowledge financial support from the Swiss National Science Foundation (to M.T.), the Office of Naval Research (to M.T., C.P.L., and A.J.H.), and the German Federal Ministry of Education and Research (to F.J.G.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5866/1066/DC1
SOM Text
Figs. S1 to S5

10 September 2007; accepted 14 January 2008
10.1126/science.1150288

Bioinspired Design and Assembly of Platelet Reinforced Polymer Films

Lorenz J. Bonderer, André R. Studart,*† Ludwig J. Gauckler†

Although strong and stiff human-made composites have long been developed, the microstructure of today's most advanced composites has yet to achieve the order and sophisticated hierarchy of hybrid materials built up by living organisms in nature. Clay-based nanocomposites with layered structure can reach notable stiffness and strength, but these properties are usually not accompanied by the ductility and flaw tolerance found in the structures generated by natural hybrid materials. By using principles found in natural composites, we showed that layered hybrid films combining high tensile strength and ductile behavior can be obtained through the bottom-up colloidal assembly of strong submicrometer-thick ceramic platelets within a ductile polymer matrix.

Steel and other metal alloys have long been used for the fabrication of strong and flaw-tolerant materials for structural applications. As opposed to metals, ceramic and polymeric materials do not exhibit the unique combination of high strength and flaw tolerance. Ceramics are typically strong but not tolerant to surface flaws and cracks, whereas most polymers are flaw-tolerant but deform extensively at rather low applied stresses.

Nature has found its way around this dilemma by combining plateletlike ceramic building blocks with polymeric matrices to render hybrid materials that are both strong and flaw-tolerant. Examples include mineralized tissues of vertebrates, such as bone, teeth, and calcified tendons, as well as the outer skeleton of invertebrates, such as the nacreous layer of mollusk shells (*1*).

The exquisite structure of these biological materials and the underlying concepts leading to their mechanical behavior have been extensively studied (*2*–*5*). Although substantial progress has been made on understanding the mechanical response of such structures (*6*–*11*), the manufacture of artificial composites that copy nature's designs remains a challenging goal.

A number of different materials have been used as inorganic reinforcing platelets for the fabrication of polymer-matrix artificial composites, including glass, graphite, SiC, AlB_2, mica, talc, and clays (*12*–*15*). Considerable increase in stiffness and strength has been achieved at rather low platelet concentrations, particularly in the case of polymers reinforced with nanosized clay particles (*16*). However, improvements in the composite strength at higher platelet concentrations [>10 weight % (wt %)] are often notably lower than that expected from theoretical models for reinforced polymers (*12*, *17*, *18*). This has been attributed to difficulties in achieving complete exfoliation and dispersion of platelets within the organic matrix and to poor bonding at the inorganic-organic interface (*12*, *18*, *19*). The strengthening seen with nanosized clay particles also seems to involve the cross-linking of polymer chains between neighboring platelets rather than solely depending on load transfer mechanisms operative in composites reinforced with micrometer-sized inorganic fibers (*17*, *18*, *20*). This cross-linking effect was used in making clay-based nanocomposites exhibiting tensile strength up to 400 MPa (*18*). However, such strong nanocomposites exhibit brittle fracture under tension (flaw intolerance), presumably because of the very high platelet concentration [50 volume % (vol %)] needed to reach this high strength level.

We applied some of the structural concepts found in biological materials to design and fabricate platelet-reinforced hybrid films that are both strong and ductile. The size and aspect ratio of the reinforcing platelets, as well as the mechanical properties of the individual inorganic and organic phases, were deliberately chosen to mimic the designing principles of strong and tough biological structures.

Biological materials such as bones, teeth, and mollusk shells are characterized by a layered structure consisting of strong inorganic platelets embedded in a soft, ductile organic matrix (*1*). In spite of the inherently weak inorganic constituents (e.g., silica, calcium carbonates, and phosphates), the high strength of the inorganic building blocks is ensured by limiting at least one of their dimensions to the nanoscale (*3*). These tiny building blocks are usually organized into a hierarchical structure spanning over various length scales. Changes in the fraction of inorganic phase (i.e., degree of mineralization) lead to hybrid materials ranging from soft tissues such as calcified tendons to strong, hard structures such as bone and nacre. Among the different models that have been proposed to describe the mechanical response of biological mineralized structures (*8*–*11*), a simple approach based on the mechanics of composite structures has been successfully applied to explain the strength and ductility of nacre (*7*).

The strength of nacre has been described on the basis of a simple shear lag model, according to which the applied load is transferred to the inorganic platelets through shear stresses developed in the organic matrix (*7*). For ductile matrices exhibiting a yield shear strength τ_y and a perfect bonding to the platelet surface (interfacial strength $\tau_i \geq \tau_y$), the ultimate tensile strength of the hybrid material (σ_c) can be estimated from the volume fraction of platelets (V_p), the platelet aspect ratio (s), and the tensile strength of the platelets (σ_p) and of the organic matrix (σ_m), as follows (*21*):

$$\sigma_c = \alpha V_p \sigma_p + (1 - V_p)\sigma_m \qquad (1)$$

with the factor α being a function of s, τ_y, and σ_p.

The tensile strength of the hybrid material depends on the operative failure mode, which in turn is determined by the aspect ratio of the inorganic platelets, s. For aspect ratios higher than a critical value ($s > s_c$), the composite fails because of the fracture of platelets (Fig. 1A), leading to a brittle catastrophic rupture of the material. In this case ($s > s_c$), the factor α in Eq. 1 is given by the following relation:

$$\alpha = 1 - \frac{\sigma_p}{2\tau_y s} \qquad (2)$$

On the other hand, for aspect ratios lower than the critical condition ($s < s_c$) the continuous matrix yields before the platelets break, thus leading to toughening mechanisms such as platelet pull-out and matrix plastic flow before the complete rupture of the composite (Fig. 1A). For $s < s_c$, the factor α reads

$$\alpha = \frac{\tau_y s}{2\sigma_p} \qquad (3)$$

Composites that fail under the platelet fracture mode are strong but brittle and thus flaw-intolerant, whereas materials that rupture under the platelet pull-out mode are generally weak but ductile and thus flaw-tolerant (Fig. 1A). Note that for weak inorganic-organic interfaces fracture occurs at the interface before the matrix

Department of Materials, Eidgenössische Technische Hochschule (ETH)–Zürich, 8093 Zürich, Switzerland.

*Present address: Experimental Soft Condensed Matter Group, Department of Physics, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.
†To whom correspondence should be addresssed. E-mail: studart@seas.harvard.edu (A.R.S.); ludwig.gauckler@mat.ethz.ch (L.J.G.)

yields, so that τ_y has to be replaced by τ_i in the composite strength Eqs. 1 and 3.

The structure of biological materials like nacre has evolved in such a way to form platelets with s ($s \sim 8$) (7) slightly below s_c, so that a maximum strength is achieved without imparting brittleness to the composite (Fig. 1B) (7). s_c is equal to the ratio σ_p/τ_y. Assuming σ_p between 360 and 500 MPa for the aragonite platelets in nacre [supporting online material (SOM) text], we expect a low s_c in the range from 9 to 12.5 for such natural composite (τ_i = 40 MPa) (7). In contrast to the limited number of inorganic materials available for the synthesis of biological structures, artificial composites can benefit from the outstanding mechanical strength of synthetic inorganic fibers and platelets ($\sigma_p \sim 2$ to 5 GPa). This leads to a higher $s_c(\sigma_p/\tau_y)$, allowing for a considerable increase of s before s_c is reached (22).

Guided by these design principles, we selected inorganic platelets and an organic polymer with

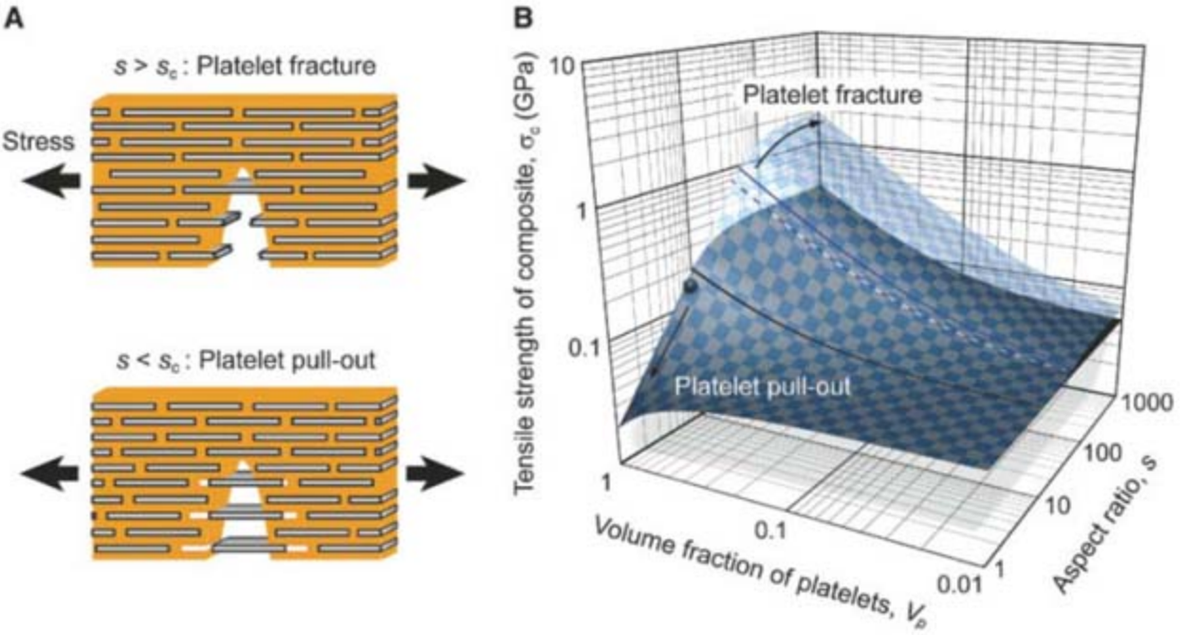


Fig. 1. Fracture mechanisms (**A**) and estimated tensile strength (**B**) of platelet-reinforced composites. The aspect ratio of the platelets (s) determines whether the composite fails under platelet fracture mode ($s > s_c$) or platelet pull-out mode [$s < s_c$ (A)]. The tensile strength of polymer matrix composites reinforced with $CaCO_3$ and Al_2O_3 platelets is estimated in (B) (black and blue surfaces, respectively) on the basis of a shear-lag mechanical model (21). The limit between the platelet fracture and pull-out modes is indicated by the black and blue full lines for $CaCO_3$ and Al_2O_3 platelets, respectively. τ_y and σ_m of, respectively, 40 and 50 MPa were assumed here for the organic matrix in nacre (protein interlamella layer) (7) and in the artificial composites (chitosan, Fig. 3B). Platelet strength (σ_p) of 0.4 and 2.0 GPa were assumed for the $CaCO_3$ (7) and Al_2O_3 platelets, respectively.

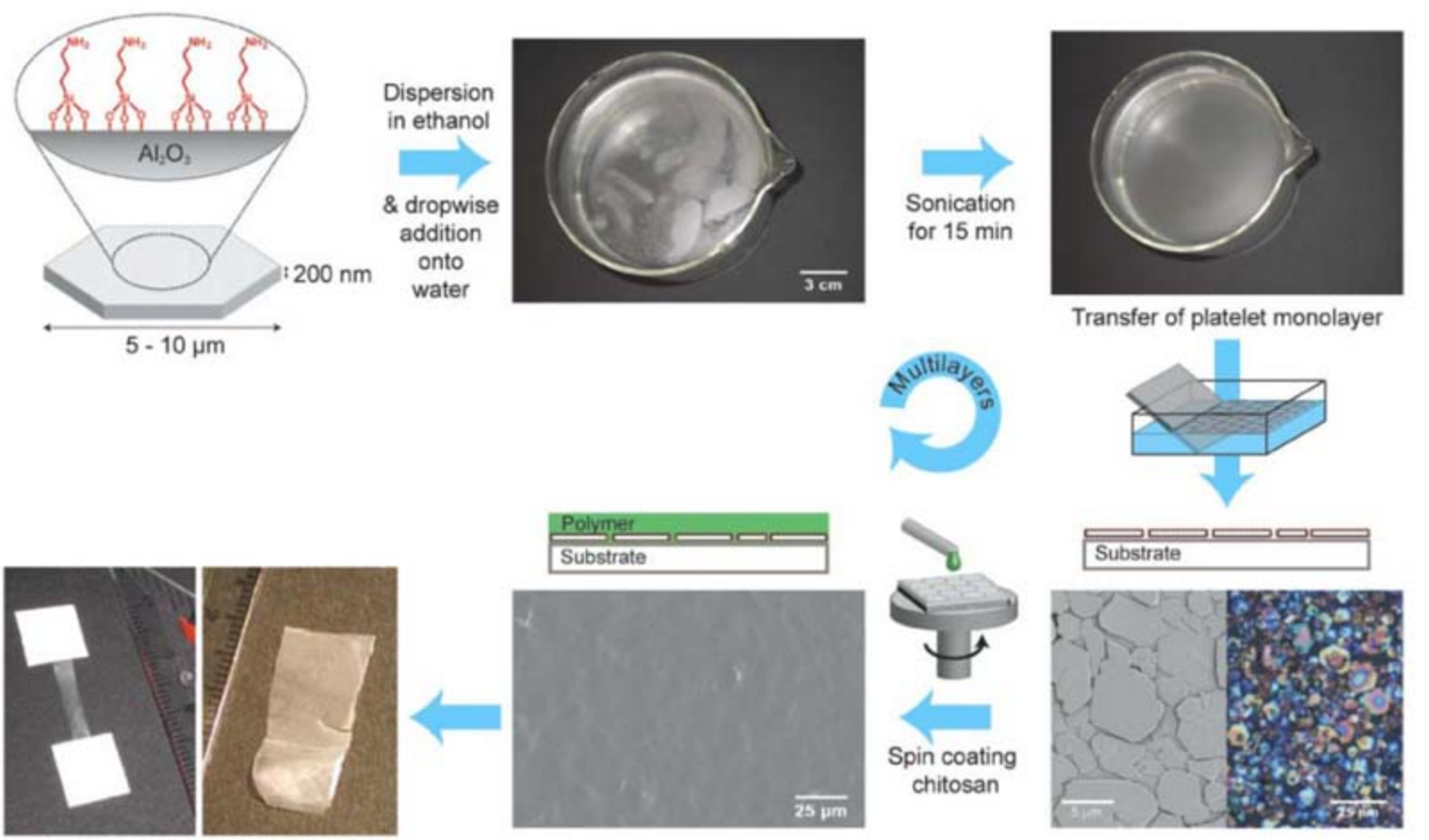


Fig. 2. Bottom-up colloidal assembly of multilayered hybrid films. Surface modified platelets are assembled at the air-water interface to produce a highly oriented layer of platelets after ultrasonication. The 2D assembled platelets are transferred to a flat substrate and afterwards covered with a polymer layer by conventional spin-coating.

features that should lead to artificial hybrid materials exhibiting high strength and ductility. Alumina platelets with estimated σ_p of 2 GPa and a chitosan polymer with τ_y around 40 MPa were chosen. With a high s_c of 50, the combination of these materials should lead to strength values higher than that of nacre while ensuring that fracture occurs by platelet pull-out, as indicated in Fig. 1B (blue surface). In order to maximize strength without impairing the polymer's ductility, 200-nm-thick platelets with an average s of about 40 were used (s slightly below s_c, see dashed blue line in Fig. 1B). In spite of their higher s, the size of the artificial alumina platelets is in the same length scale as that of the aragonite platelets encountered in nacre (*3*, *7*). With the exception of a recent study on composites with low platelet concentrations ($V_p \leq 0.03$) (*15*), most investigations have been carried out with use of either silicate-based platelets with thickness around 1 nm (*13*, *14*, *16*–*18*, *20*, *23*, *24*) or artificial platelets with thickness larger than a few tens of a micrometer (*19*, *22*). Platelets at the length scale we used are expected to be thin enough to possess the high tensile strength of small inorganic building blocks (*3*) and sufficiently thick to be fully dispersed in the polymer matrix so as to strengthen the composite through conventional load transfer mechanisms (*22*).

We fabricated layered hybrid materials by using an approach that relies on the sequential deposition of inorganic and organic layers at ambient conditions (Fig. 2). Because the artificial platelets are submicrometer-sized in thickness, colloidal-based techniques can be used for the assembly of the inorganic building blocks. The ability of colloidal platelets to adsorb at air-water interfaces was used as a means to direct their assembly into a highly oriented two-dimensional (2D) structure. The adsorption of platelets at the air-water interface was favored through the attachment of slightly hydrophobic amine-terminated silane species on the platelet surface (fig. S1) (*25*). Amine groups at the end of the silane hydrophobic tail are expected to form hydrogen bonds with the oxygen atoms of the chitosan backbone, increasing the adhesion between inorganic platelets and the organic matrix. An ethanol suspension containing 1 vol % of modified platelets was spread over a water surface. A smooth and perfectly oriented monolayer of platelets was formed at the surface of water upon ultrasonication (Fig. 2). The formation of a repulsive electrical double layer between like-charged platelets at neutral pH allowed for extensive rearrangement of particles during sonication. The 2D assembled platelets were transferred to a glass substrate by dip-coating and were then spin-coated with an organic layer of chitosan solution (*25*). The thickness of the polymer layer was controlled by changing the chitosan concentration in the spin-coating solution (fig. S2). Repetition of these steps in a sequential manner leads to multilayered inorganic-organic films with a total thickness typically less than a few tens of a micrometer. Free-standing films were obtained by peeling them off from the substrate with a razor blade (Fig. 2).

Flexible and thin hybrid films containing V_p up to 0.2 exhibited a brick-mortar structure with strongly aligned platelets surrounded by a ductile organic matrix (Fig. 3A). For inorganic volume fractions higher than 0.2, extensive swelling of the thin chitosan layer during dip coating led to platelet misalignment and the incorporation of voids within the film (Fig. 4A and fig. S3). As a result of strong platelet alignment for volume fractions lower than 0.2 (Fig. 4A), a substantial fraction of a load applied parallel to the ordered layers can be taken by the stronger inorganic phase, increasing the material's elastic modulus (E_c) from 2 up to 10 GPa (Figs. 3B and 4B).

The mechanical behavior of the hybrid film strongly deviated from the linear elastic regime, when the yield tensile strength of the organic matrix was reached. At this stress condition, yielding of the polymer phase between the inorganic platelets led to a pronounced plastic deformation of the composite film (Fig. 3B). Because of load transfer to the platelets, the tensile stresses required for plastic yielding increased from 50 MPa to values as high as 300 MPa when the V_p was increased from 0 to 0.15 (Figs. 3B and 4C). Most remarkably, films containing inorganic volume fractions up to 0.15 fractured at a total strain (ε_{rupt}) typically between 4 and 35% (Fig. 4D), as a result of extensive plastic yielding of the polymeric matrix before rupture. Flexible hybrid films that are simultaneously strong (tensile strength $\sigma_c \sim 300$ MPa) and ductile ($\varepsilon_{rupt} \sim 20\%$) were

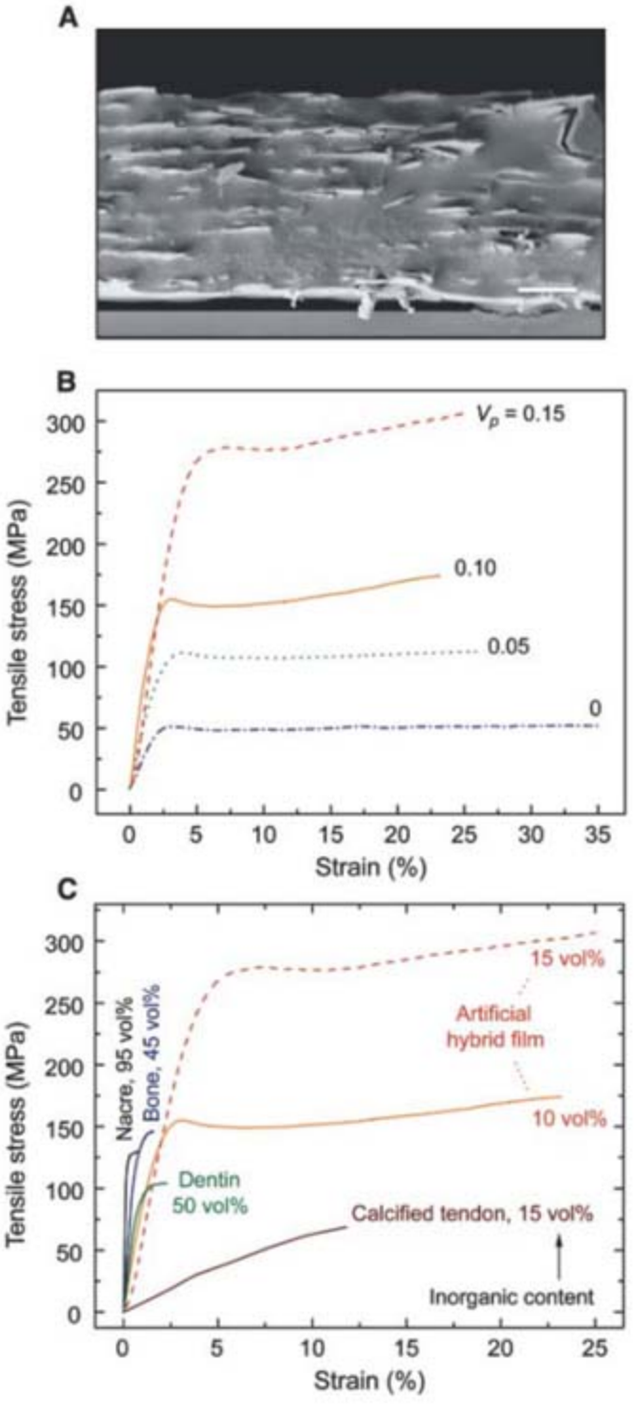


Fig. 3. Microstructure and mechanical behavior of artificial and natural hybrid materials with different concentrations of inorganic platelets. The highly oriented microstructure of an Al_2O_3-chitosan artificial film is shown in (**A**) for V_p = 0.11. The use of strong platelets with s slightly lower than s_c leads to light and flexible composites combining high tensile strength and high strain at rupture (**B**). Chitosan-based films prepared with platelet concentrations as low as 10 and 15 vol % exhibit tensile strength comparable to or higher than that of natural layered materials (**C**). The stress versus strain curves for nacre (red abalone, *Haliotis rufescens*), bone, dentin, and calcified tendon were obtained from Wang *et al.* (*2*), Landis *et al.* (after Ashman) (*29*), Sano *et al.* (*28*), and Landis *et al.* (*29*), respectively. The scale bar in (A) indicates 5 μm. The stress versus strain curves shown in (B) and (C) are representative curves of the total of five evaluated samples.

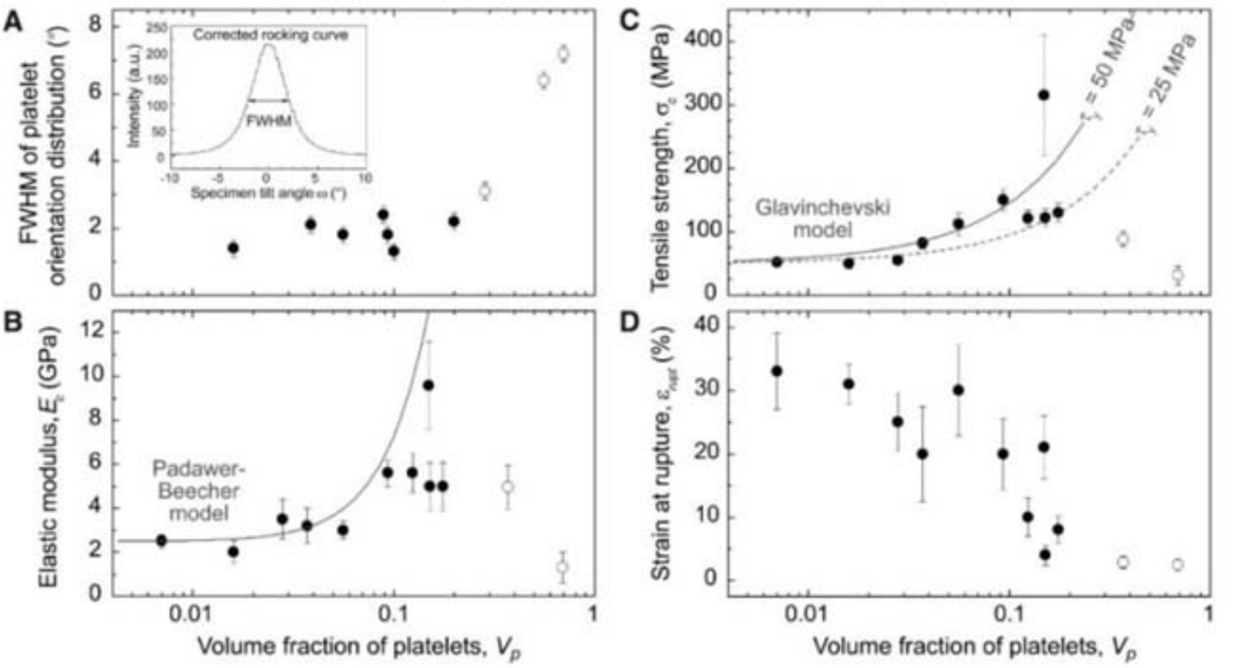


Fig. 4. Platelet orientation and mechanical properties of Al_2O_3-chitosan artificial films as a function of V_p. A high degree of orientation was achieved for values of V_p up to 0.2 [(**A**), solid symbols]. Load transfer from the continuous chitosan phase to the highly oriented inorganic platelets led to a substantial increase in E_c (**B**) and σ_c of the hybrid film (**C**), in accordance with mechanical models for bulk composite materials [continuous fitting curves in (B) and (C)] (*21*, *27*). Although the strain at rupture of the hybrid films decreases by increasing the concentration of platelets, values as high as 4% are still achieved for V_p = 0.2 (**D**). The elastic moduli data (B) were obtained from the initial elastic region of the stress versus strain curves exemplified in Fig. 3B. Error bars correspond to the standard deviation obtained from a total of five samples.

successfully produced at rather low V_p values (V_p = 0.15) (Fig. 3B).

Our results indicate that the interfacial bonding between platelets and the organic matrix plays a crucial role on the load transfer efficiency from the ductile polymer to the strong inorganic phase. For τ_i values lower than the matrix τ_y, the ultimate strength of the composite is controlled by τ_i rather than τ_y (Eqs. 1 and 3). A comparison of our experimental results with those expected from the simplified shear-lag model (Fig. 4C) shows that most of the strength data agree well with the theoretical predictions when τ_i is varied from 25 to 50 MPa (*21*). Films exhibiting the highest σ_c (σ_c ~ 300 MPa) are stronger than the model predictions, suggesting that in this case τ_i or τ_y might have been improved during the processing of the hybrid film. A possible reason for an increase of the τ_y would be the crystallization of chitosan molecules on the surface of the alumina platelets.

The E_c data for the layered composites are also in reasonable agreement with theoretical predictions (*26*, *27*) (Fig. 4B), except for some data sets at V_p between 0.1 and 0.2. Scattering of the strength and elastic modulus at this range of platelet concentration (Fig. 4, B and C) was probably caused by the limited control over the organic-inorganic interface of the Al_2O_3-chitosan films or by the possible presence of processing defects (voids) within the composite microstructure. Improvements in the sequential assembly process to avoid processing defects and a better control over the interfacial bonding by chemically coupling the platelet surface groups to the organic matrix should considerably reduce scattering of the strength and E_c values.

The combination of high tensile strength and ductility (inelastic strain) is the most distinct feature of our composites when compared to other platelet-reinforced polymers (Table 1) (*14*, *18*, *20*, *23*). Composites with 15 vol % of submicrometer alumina platelets showed tensile strength up to 315 MPa and an inelastic deformation of 17%, as opposed to the catastrophic brittle failure observed for nanocomposites containing up to 50 vol % of clay platelets (Table 1) (*14*, *18*). Such inelastic deformation should lead to very high fracture energy and toughness. A rough estimation based on the area under the stress versus strain curve indicates that the energy required to rupture our composites is two orders of magnitude higher than that needed to fracture ultrastrong and stiff clay-based nanocomposites (*14*, *18*). The lower platelet concentration of the alumina-reinforced polymer led to relative strength (σ_c/σ_m) and E_c values that are, respectively, 40% and 10-fold lower than that of ultrastrong nanocomposites (Table 1) (*18*). It is important to note, however, that the submicrometer-thick platelets used exhibit a higher reinforcing efficiency compared with the clay platelets investigated by Podsiadlo *et al.* (*18*). On the basis of the factor α of Eq. 1, we estimated that the submicrometer alumina platelets are at least four times more efficient than clay platelets as load-bearing units of the composite (SOM text).

Table 1. Summary of the mechanical properties of artificial composites reinforced with submicrometer-thick alumina platelets in comparison to clay-based nanocomposites (*14*, *18*). Data refer to mean values and corresponding standard deviation. *N* indicates the minimum number of samples evaluated. PDDA, poly(diallyldimethylammonium chloride); PVA, poly(vinyl alcohol).

Composite (*N*)	V_p	σ_c (MPa)	Relative strength (σ_c/σ_m)	E_c (GPa)	ε_{rupt} (%)	Elastic strain, ε_{el} (%)	Inelastic strain, ε_{inel} (%)
Submicrometer Al_2O_3 platelet/chitosan (*5*)	0.15	315 ± 95	6.0	9.6 ± 2	21 ± 5	3.9 ± 1.2	17.1 ± 4
Montmorillonite clay/PDDA	~0.30*	100 ± 10	8.3	11 ± 2	10 ± 2†	0.9*	0
Montmorillonite clay/cross-linked PVA	0.50	400 ± 40	10.0	106 ± 11	0.33 ± 0.04	0.33 ± 0.04	0

*Estimated based on data reported by Tang *et al.* (*14*). †According to data published by Tang *et al.* (*14*), these samples only bear considerable load for strain values higher than 9%.

The mechanical behavior of the artificial composites was also compared with that of hybrid biological structures in Fig. 3C. In spite of their lower inorganic content, the artificial composites exhibit yield strength higher than or comparable to that of natural materials. As anticipated by the mechanical model shown in Fig. 1, the high strength of artificial composites prepared at rather low inorganic contents results from the enhanced tensile strength of the synthetic inorganic platelets (σ_p ~ 2 GPa) as compared to that of platelets found in biological structures [360 < σ_p < 500 MPa in nacre (SOM text)]. The strain to rupture of the artificial composites is also considerably higher than that of natural materials because of the extensive plastic flow of the chitosan matrix (Fig. 3C). The area under the stress versus strain curves obtained for the artificial materials is more than one order of magnitude larger than that for

nacre and other biological materials (Fig. 3C). The elastic moduli of up to 10 GPa achieved by the artificial composite is comparable to those of dentin and bone (*9*, *28*, *29*) and about 10-fold lower than that of nacre (*7*).

The use of high-strength artificial platelets as mechanical reinforcement imposes less-stringent microstructural requirements to the composites in comparison to natural hybrid structures. Artificial composites reinforced with strong platelets show remarkable mechanical properties in spite of their less-elaborate microstructure. In contrast, the relatively weak inorganic compounds available in nature for biomineralization require a far more sophisticated architecture to render materials with comparable mechanical behavior. Nature still remains supreme in its ability to build hybrid materials with unique structures and properties using a relatively limited variety of inorganic building blocks. Further advances in this area might allow us to replicate the microstructure of natural materials by using strong artificial building blocks in the future. If successful, this bio-inspired approach would eventually lead to man-made hybrid materials with unprecedented mechanical properties.

References and Notes

1. H. A. Lowenstam, S. Weiner, *On Biomineralization* (Oxford Univ. Press, New York, 1989), pp. 336.
2. R. Z. Wang, Z. Suo, A. G. Evans, N. Yao, I. A. Aksay, *J. Mater. Res.* **16**, 2485 (2001).
3. H. J. Gao, B. H. Ji, I. L. Jager, E. Arzt, P. Fratzl, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5597 (2003).
4. G. Mayer, *Science* **310**, 1144 (2005).
5. N. Nassif *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12653 (2005).
6. J. D. Currey, *Proc. R. Soc. London Ser. B* **196**, 443 (1977).
7. A. P. Jackson, J. F. V. Vincent, R. M. Turner, *Proc. R. Soc. London Ser. B* **234**, 415 (1988).
8. H. D. Wagner, S. Weiner, *J. Biomech.* **25**, 1311 (1992).
9. I. Jager, P. Fratzl, *Biophys. J.* **79**, 1737 (2000).
10. A. G. Evans *et al.*, *J. Mater. Res.* **16**, 2475 (2001).
11. F. Barthelat, H. Tang, P. D. Zavattieri, C. M. Li, H. D. Espinosa, *J. Mech. Phys. Solids* **55**, 306 (2007).
12. J. Rexer, E. Anderson, *Polym. Eng. Sci.* **19**, 1 (1979).
13. E. R. Kleinfeld, G. S. Ferguson, *Science* **265**, 370 (1994).
14. Z. Y. Tang, N. A. Kotov, S. Magonov, B. Ozturk, *Nat. Mater.* **2**, 413 (2003).
15. I. Chasiotis, Q. Chen, G. M. Odegard, T. S. Gates, *Exp. Mech.* **45**, 507 (2005).
16. P. B. Messersmith, E. P. Giannelis, *Chem. Mater.* **6**, 1719 (1994).
17. D. Schmidt, D. Shah, E. P. Giannelis, *Curr. Opin. Solid State Mater. Sci.* **6**, 205 (2002).
18. P. Podsiadlo *et al.*, *Science* **318**, 80 (2007).
19. N. Sheng *et al.*, *Polymers* **45**, 487 (2004).
20. A. Okada, A. Usuki, *Mater. Sci. Eng. C* **3**, 109 (1995).
21. B. Glavinchevski, M. Piggott, *J. Mater. Sci.* **8**, 1373 (1973).
22. D. Hull, T. W. Clyne, *An Introduction to Composite Materials*, D. R. Clark, S. Suresh, I. M. Ward, Eds., Cambridge Solid State Science Series (Cambridge Univ. Press, Cambridge, ed. 2, 1996).
23. D. G. H. Ballard, G. R. Rideal, *J. Mater. Sci.* **18**, 545 (1983).
24. N. Almqvist *et al.*, *Mater. Sci. Eng. C* **7**, 37 (1999).
25. Materials and methods are available on *Science* Online.
26. H. L. Cox, *Br. J. Appl. Phys.* **3**, 72 (1952).
27. G. E. Padawer, N. Beecher, *Polym. Eng. Sci.* **10**, 185 (1970).
28. H. Sano, B. Ciucchi, W. G. Matthews, D. H. Pashley, *J. Dent. Res.* **73**, 1205 (1994).
29. W. J. Landis, J. J. Librizzi, M. G. Dunn, F. H. Silver, *J. Bone Miner. Res.* **10**, 859 (1995).

30. We thank V. Klass, I. Olliges, and C. Müller for their contribution to the experimental part of this study and ETH Zurich for financial support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1069/DC1
Materials and Methods
SOM Text
Figs. S1 to S3

2 August 2007; accepted 4 January 2008
10.1126/science.1148726

Atomic-Scale Chemical Imaging of Composition and Bonding by Aberration-Corrected Microscopy

D. A. Muller,[1,2]* L. Fitting Kourkoutis,[1] M. Murfitt,[3] J. H. Song,[4,5] H. Y. Hwang,[5,6] J. Silcox,[1,2] N. Dellby,[3] O. L. Krivanek[3]

Using a fifth-order aberration-corrected scanning transmission electron microscope, which provides a factor of 100 increase in signal over an uncorrected instrument, we demonstrated two-dimensional elemental and valence-sensitive imaging at atomic resolution by means of electron energy-loss spectroscopy, with acquisition times of well under a minute (for a 4096-pixel image). Applying this method to the study of a $La_{0.7}Sr_{0.3}MnO_3/SrTiO_3$ multilayer, we found an asymmetry between the chemical intermixing on the manganese-titanium and lanthanum-strontium sublattices. The measured changes in the titanium bonding as the local environment changed allowed us to distinguish chemical interdiffusion from imaging artifacts.

Resolving and identifying the composition and bonding of all atoms in a solid sample at atomic resolution has been a long-standing goal of analytical microscopy. Atom probe microscopy comes close to the first goal of compositional identification, collecting roughly 60% of the atoms field-evaporated off a sharp tip and locating their positions to within a few lattice sites. Electron energy-loss spectroscopy (EELS) of a high-energy electron beam focused down to atomic dimensions allows elemental identification with atomic resolution, with the additional capabilities of imaging insulating materials and determining the electronic structure or chemical bonding of each atomic column in the material (*1*).

One step toward atomic-resolution EELS imaging was the identification of the chemical state at a single point at an interface, with the use of a scanning transmission electron microscope (STEM) (*2*). However, the resolution was limited by a small incident electron beam current, allowing bonding states to be mapped at subnanometer but not atomic resolution (*3*). The recent correction of third-order electron-optical aberrations by multipole correctors has allowed the illumination aperture size to be increased, yielding a factor of 4 to 6 increase in the beam current for the same probe size (*4*). This has made it possible to record atomic-resolution EELS line profiles (*5*) or chemical images without bonding information (*6*, *7*).

A full two-dimensional chemical map obtained by EELS that also contains bonding information requires yet another order of magnitude increase in the beam current to improve the signal-to-noise ratio. We did so by correcting the electron optical aberrations to fifth-order so the numerical aperture in the STEM could be further increased (*8*). However, even this increase in beam current was not sufficient to produce a true chemical map. Electrons are scattered both elastically and inelastically in the sample. Although the inelastic scattering that contains the EELS signal is peaked largely in the forward direction, the unavoidable elastic scattering of those same electrons is strongly angle-dependent. Standard collection optics capture only a portion of the inelastically scattered electrons, because heavier atoms elastically scatter them out of the path to the detector. Consequently, the EELS signal can be dominated by changes in the elastic contrast, masking the chemically unique inelastic signal.

This phenomenon was recently demonstrated, both by experiment and quantitative theory, for $Bi_{0.5}Sr_{0.5}MnO_3$, where the "inelastic" images were dominated by the strong elastic scattering from the Bi sites that largely prevented electrons on the Bi column from entering the detector (*6*). A general symptom of this scattering artifact is that all inelastic intensities, independent of their characteristic energy loss, would peak at the same spatial locations [as was mentioned in (*9*)]. The artifact can be eliminated if all the elastically scattered electrons are being collected by the spectrometer. However, increasing the collection angle of a conventional spectrometer to do this is not possible without drastically degrading the

[1]Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA. [2]Kavli Institute at Cornell, Cornell University, Ithaca, NY 14853, USA. [3]Nion Co., Kirkland, WA 98033, USA. [4]Department of Physics, Chungnam National University, Daejeon 305-764, Korea. [5]Department of Advanced Materials Science, University of Tokyo, Kashiwa, Chiba 277-8561, Japan. [6]Japan Science and Technology Agency, Kawaguchi 332-0012, Japan.

*To whom correspondence should be addressed. E-mail: dm24@cornell.edu

energy resolution to the point that the EELS edges are no longer detectable (*10*). By adding four round coupling lenses that parallelize the scattered electrons, plus a corrective quadrupole/octupole module before the spectrometer (*8*), we obtained 0.5-eV energy resolution for collection angles as large as 60 mrad. These angles are larger than those in a conventional system by a factor of roughly 3 to 6, and they collect a much larger fraction of the elastically scattered energy-loss electrons and thereby reduce the measured elastic modulation to below a few percent.

Figure 1 shows atomic-resolution EELS maps that demonstrate this capability. The underlying spectrum image was recorded with an incident beam current of ~780 pA, with ~600 pA collected in the spectrometer (*11*). For comparison, the reported incident beam currents of third-order correctors range from 10 to 100 pA when used for EELS mapping, with an estimated usable EELS signal of 8 to 50 pA (*11*). This factor of 10 to 100 increase in signal can be attributed to a high-brightness source, improved probe formation, and efficient signal collection. An uncorrected microscope has a probe-forming aperture semi-angle of ~10 mrad. A third-order corrector allows this angle to be doubled, increasing the beam current fourfold for the same source size. A fifth-order corrector allows the angle to be doubled yet again, giving a factor of 16 increase in beam current over the uncorrected system. An additional factor of 10 in current is gained by using a cold field emitter instead of the more common Schottky emitter. Finally, the post-specimen optics were matched to collect over a much wider range of scattering angles than in uncorrected instruments (*8*).

The sample we studied is a perovskite $La_{0.7}Sr_{0.3}MnO_3/SrTiO_3$ (LSMO/STO) multilayer, with five nominal unit cells of each material, grown in conditions where cation interdiffusion occurs; such a multilayer is useful for the bonding studies presented below (*11*). This sample is a good resolution test because there are two distinct sublattices, Sr-La and Mn-Ti, offset from each other by 0.19 nm, and the edges of the multilayers themselves provide a "knife-edge" resolution test where the resulting image is essentially a convolution of the probe shape with a step function. This is a useful check for probe tails that are filtered out of a periodic lattice image. Both tests are needed to distinguish between true inelastic contrast and the preservation of the elastic contrast in the inelastic channel (*12*). Here, the La and Mn lattice images are properly offset from each other (Fig. 1, A and C), showing that the elastic contrast artifact (which would cause them to coincide) has been suppressed.

Figure 2A shows some of the individual spectra that make up the chemical map, and the simultaneously recorded annular dark-field image from the elastic signal (Fig. 2B) shows that the EELS signals are properly registered with their respective atomic columns. The sharp left edge of the STO layer visible as the dark band in the center of Fig. 1A shows the abrupt profile in the La signal. Another map of a nearby region of the sample, recorded with the same probe alignments, shows an abrupt edge profile for Ti (*11*) (fig. S2). This counterexample demonstrates that the more gradual interfaces for the Ti and Mn signals in Fig. 1, B and C, are not the result of inelastic delocalization [which will be worst for Ti, the lowest-energy edge (*13*–*15*)] or of probe spreading, but rather are a property of the sample itself.

With these data we were able to explore chemical bonding states at an atomic scale. The shape of the EELS edge reflects the local, final-state density of states modified by the presence of a core hole on a specific atom column (*16*). The bonding changes are most noticeable at an energy scale of 0.2 to 2 eV, so a simple check that we have sufficient signal to image bonding states is to display maps of the EELS fine structure formed with energy windows in this range (Fig. 3, A and B, for Ti and Mn, respectively, using a 0.5-eV window). The changes in EELS fine structure are complex, but some general bonding trends are captured by a principal components analysis of the spectrum image (*17*). The component that contains the largest image variance, PC1 (Fig. 3C), tracks the average shape of the Ti L edge and follows the map of total Ti signal. We found that the second-largest component, PC2 (Fig. 3D), tracks the differences in Ti bonding between Ti-rich STO and Ti-poor LSMO layers (spectra shown in Fig. 3E). PC2 is only ~15% the intensity of PC1, so the lattice fringe modulation is reduced below the noise level in PC2. Nonetheless, the presence of PC2 allowed us to distinguish between imaging artifacts, such as the small but unavoidable probe tails due to beam spreading, and chemical interdiffusion in the sample itself. Probe tails will blur the Ti signal of PC1 spatially but

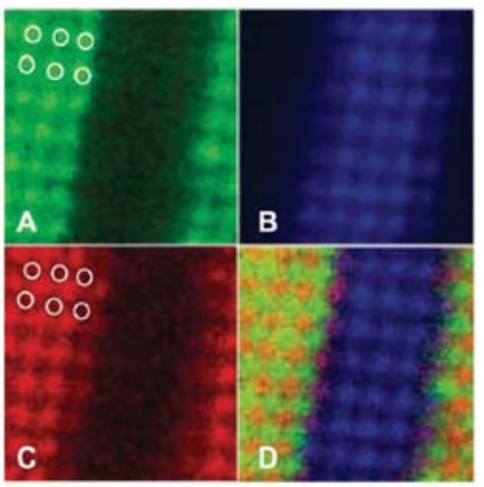


Fig. 1. Spectroscopic imaging of a $La_{0.7}Sr_{0.3}MnO_3/SrTiO_3$ multilayer, showing the different chemical sublattices in a 64 × 64 pixel spectrum image extracted from 650 eV–wide electron energy-loss spectra recorded at each pixel. (**A**) La M edge; (**B**) Ti L edge; (**C**) Mn L edge; (**D**) red-green-blue false-color image obtained by combining the rescaled Mn, La, and Ti images. Each of the primary color maps is rescaled to include all data points within two standard deviations of the image mean. Note the lines of purple at the interface in (D), which indicate Mn-Ti intermixing on the B-site sublattice. The white circles indicate the position of the La columns, showing that the Mn lattice is offset. Live acquisition time for the 64 × 64 spectrum image was ~30 s; field of view, 3.1 nm.

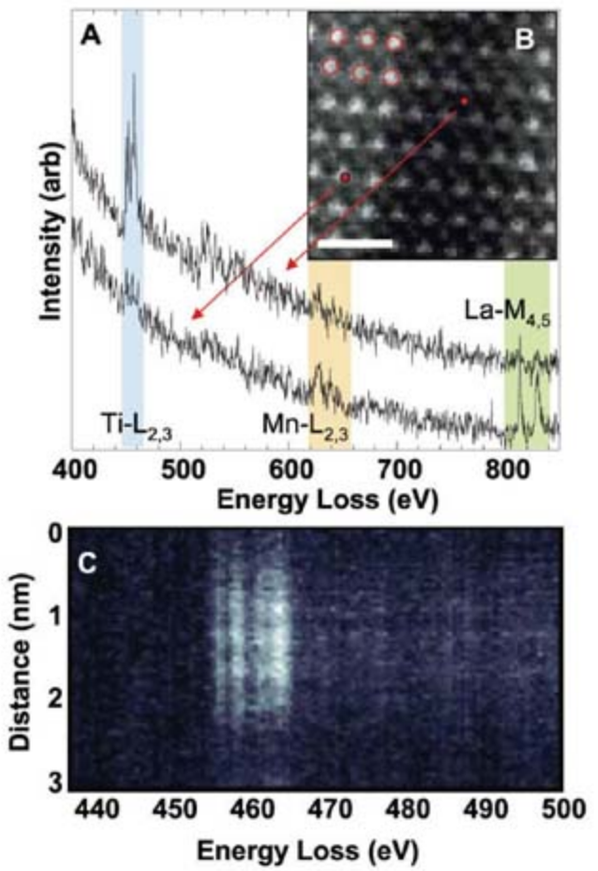


Fig. 2. (**A**) Individual EELS spectra from the series shown in Fig. 1. (**B**) The simultaneously recorded annular dark-field (ADF) image. The large red circles show that the La signal from Fig. 1A is correctly peaked where the ADF scattering is strongest. The small red dots indicate the spatial locations from which the EELS spectra were selected. Scale bar, 1 nm. (**C**) Background-subtracted and gain-corrected Ti L-edge spectra, plotted for a single line across the $SrTiO_3$ layer. The four fine-structure peaks corresponding to the unoccupied Ti 3d e_g and t_{2g} levels for the Ti L_3 and Ti L_2 edges are resolved.

cannot change the shape of the spectrum, as PC2 does. Instead, whereas the Ti in the STO layer shows EELS fine structure typical of bulk $SrTiO_3$, the Ti in the LSMO layer shows a strong damping of the crystal-field split levels, as is typical for local disorder such as binding to vacancies or other defects (*18*, *19*). Bonding changes on the oxygen sublattice are shown in fig. S3, again using a principal components analysis that is found to separate total concentration from chemical changes (*11*). Additional checks on common artifacts from beam spreading and delocalization are also given in (*11*).

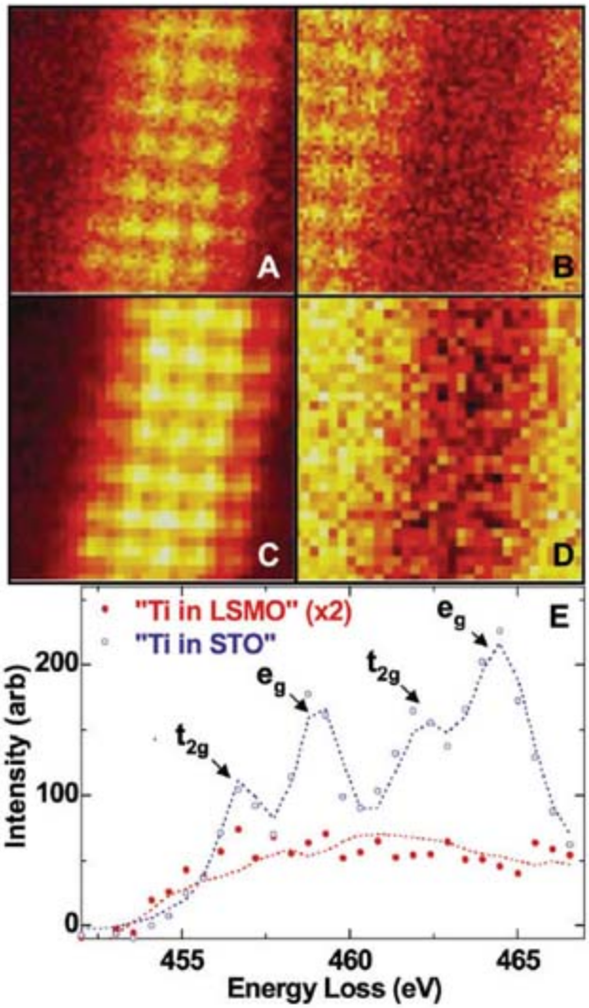


Fig. 3. Two-dimensional mapping of bonding details for the Ti and Mn 3d states, using only a single 0.5 eV–wide channel in the spectrum image (**A**) at the peak of the Ti L_2 e_g EELS edge, and (**B**) at the peak of the Mn L_3 edge, showing sufficient signal to form atomic-resolution images of the final-state density of states. To explore bonding variations across the multilayer, the Ti L spectrum image is rebinned to 32 × 32 pixels to improve the signal-to-noise ratio. (**C** and **D**) A principal components analysis gives two statistically significant components to the edge. The intensity scale of the second component is roughly 15% of that of the first component, reflecting the reduction in Ti concentration in the LSMO layer where this component is peaked. (**E**) The Ti L edge showing the crystal-field split Ti 3d e_g and t_{2g} empty states. The leftmost two peaks are transitions from the L_3 core level, and the rightmost two peaks are from the L_2 core level. The open and solid points are from single points (after rebinning) in the Ti-rich STO layer ("Ti in STO") and the Ti-poor LSMO layer ("Ti in LSMO"). The dotted lines are the fits from the first two principal components shown in (C) and (D). Field of view for all images, 3.1 nm.

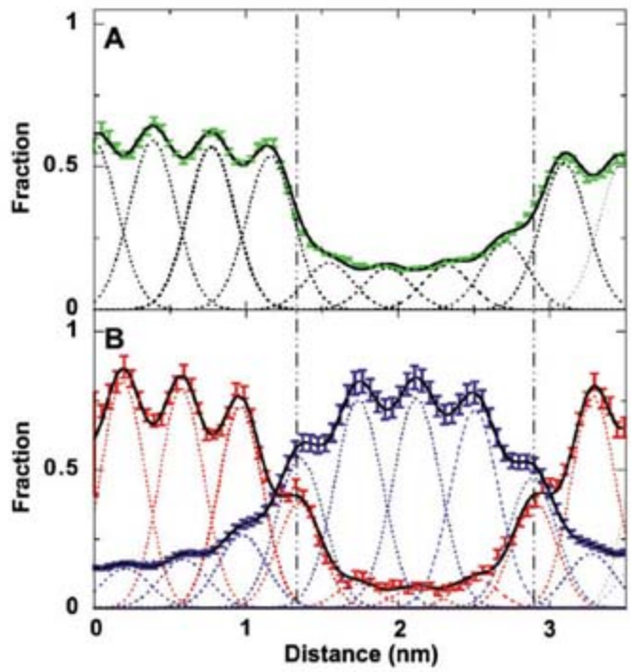


Fig. 4. Line profiles averaged across the least-squares fit to the concentration maps of Fig. 1, corrected for the drift during acquisition. (**A**) The La M-edge concentration profile shows a small amount of La present inside the nominal $SrTiO_3$ layer and an asymmetry in intermixing between the sharp lower (left) and more gradual upper (right) interfaces. The black curve is a fit to a sum of Gaussian peaks fixed at the lattice sites (dashed), which gives an indication of the local La concentration partitioned by plane, showing the nominal $SrTiO_3$ is four unit cells thick in this region of the sample. (**B**) The Mn (red) and Ti (blue) profiles show a more symmetric intermixing at the two interfaces. The Ti displaced from these planes is spread throughout the LSMO layer, but the Mn in the STO layer is largely confined to these outermost planes. For all three curves, the colored error bars shown are the 95% confidence interval for the standard error of the mean (±2 SEM) calculated from the 64 different line profiles that make up each map.

Using one-component least-squares fits, we also built up concentration maps for La and Mn. These do not look significantly different from the simple summations of Fig. 1. The statistics for any given spectrum are relatively poor, but by averaging across the layer, the random fluctuations can be averaged out and we can obtain unbiased estimates of the random errors in the concentration measurement that we would not be able to obtain from a single line scan. Figure 4A shows that significant La has been incorporated in the nominal STO layer. The La profile at the left interface is abrupt, but at the right interface it is more gradual. There is also significant intermixing on the Mn-Ti sublattice at the last atomic plane in the STO layer (Fig. 4B). The chemical changes on the Ti edges show that these are different atoms, rather than probe tails from the neighboring layer (Fig. 3). (The presence of the core-level shift rules out the possibility that this is an inelastic tail from the STO layer, which would maintain the same shape.) The local concentration profiles are extracted by fitting a series of Gaussian peaks at each of the atomic columns. The width of the Gaussian reflects the averaging over random scan noise fluctuation between the 64 different image lines, rather than the actual and much smaller probe profile itself.

The experimental data show an asymmetry between A-site (La-Sr) and B-site (Mn-Ti) intermixing. Because the thin-film behavior of manganites shows a critical thickness behavior for conductivity, although the values vary (*20*–*22*), local compositional changes such as those seen here may have a key role to play.

We have demonstrated two-dimensional chemical imaging at the atomic scale, with simultaneous elemental identification and visualization of the local bonding states. The corrected optics allowed us to increase the convergence and collection angles. As a consequence, the elastic scattering artifacts that masked the chemical signal in earlier EELS work have been suppressed, and the image acquisition times were reduced by a factor of 10 to 100.

References and Notes

1. J. C. H. Spence, J. Lynch, *Ultramicroscopy* **9**, 267 (1982).
2. P. E. Batson, *Nature* **366**, 727 (1993).
3. D. A. Muller, Y. Tzou, R. Raj, J. Silcox, *Nature* **366**, 725 (1993).
4. P. E. Batson, N. Dellby, O. L. Krivanek, *Nature* **418**, 617 (2002).
5. L. J. Allen, S. D. Findlay, A. R. Lupini, M. P. Oxley, S. J. Pennycook, *Phys. Rev. Lett.* **91**, 105503 (2003).
6. M. Bosman *et al.*, *Phys. Rev. Lett.* **99**, 086102 (2007).
7. K. Kimoto *et al.*, *Nature* **450**, 702 (2007).
8. O. L. Krivanek *et al.*, *Ultramicroscopy*, 10.1016/j.ultramic.2007.07.010 (2007).
9. E. Okunishi, H. Sawada, Y. Kondo, M. Kersker, *Microsc. Microanal.* **12** (suppl. 2), 1150 (2006).
10. H. A. Brink, M. M. G. Barfels, R. P. Burgner, B. N. Edwards, *Ultramicroscopy* **96**, 367 (2003).
11. See supporting material on *Science* Online.

12. P. Stallknecht, H. Kohl, *Ultramicroscopy* **66**, 261 (1996).
13. E. C. Cosgriff, M. P. Oxley, L. J. Allen, S. J. Pennycook, *Ultramicroscopy* **102**, 317 (2005).
14. H. Kohl, H. Rose, *Adv. Imaging. Electron Phys.* **65**, 173 (1985).
15. D. A. Muller, J. Silcox, *Ultramicroscopy* **59**, 195 (1995).
16. R. F. Egerton, *Electron Energy Loss Spectroscopy in the Electron Microscope* (Plenum, New York, ed. 2, 1996).
17. P. Trebbia, N. Bonnet, *Ultramicroscopy* **34**, 165 (1990).
18. J. G. Chen, *Surf. Sci. Rep.* **30**, 1 (1997).
19. D. A. Muller, N. Nakagawa, A. Ohtomo, J. L. Grazul, H. Y. Hwang, *Nature* **430**, 657 (2004).
20. K. Dorr *et al.*, *J. Appl. Phys.* **89**, 6973 (2001).
21. X. Hong, A. Posadas, C. H. Ahn, *Appl. Phys. Lett.* **86**, 142501 (2005).
22. M. Izumi *et al.*, *Phys. Rev. B* **64**, 064429 (2001).
23. The instrument was developed under NSF grant DMR-9977547 with an additional contribution from Cornell University. In-kind contributions were made by Gatan Inc., Sandia National Laboratory, Pacific Northwest Laboratory, and Bell Laboratories. Supported by NSF Nanoscale Science and Engineering Initiative grant EEC-0117770 (D.A.M. and J.S.), NSF Materials Research Science and Engineering Center grant DMR-0520404 (L.F.K.), the TEPCO Research Foundation (H.Y.H.), and the Japan Society for the Promotion of Science (J.H.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5866/1073/DC1
Materials and Methods
Figs. S1 to S3
References

6 August 2007; accepted 18 January 2008
10.1126/science.1148820

Extending Earthquakes' Reach Through Cascading

David Marsan* and Olivier Lengliné

Earthquakes, whatever their size, can trigger other earthquakes. Mainshocks cause aftershocks to occur, which in turn activate their own local aftershock sequences, resulting in a cascade of triggering that extends the reach of the initial mainshock. A long-lasting difficulty is to determine which earthquakes are connected, either directly or indirectly. Here we show that this causal structure can be found probabilistically, with no a priori model nor parameterization. Large regional earthquakes are found to have a short direct influence in comparison to the overall aftershock sequence duration. Relative to these large mainshocks, small earthquakes collectively have a greater effect on triggering. Hence, cascade triggering is a key component in earthquake interactions.

Earthquakes of all sizes, including aftershocks, are able to trigger their own aftershocks. The cascade of earthquake triggering causes the seismicity to develop complex, scale-invariant patterns. The causality of "mainshock A triggered aftershock B," which appears so obvious if mainshock A happens to be large, must then be modified into a more subtle "mainshock A triggered C1, which triggered C2, ..., which triggered B." This has paramount consequences: The physical mechanism that causes direct triggering (static or dynamic stress changes, fluid flow, afterslip, etc.) cannot be studied by looking at aftershocks that were not directly triggered by the mainshock. Moreover, if indirect triggering is important in the overall aftershock budget (*1*–*3*), then direct triggering must be confined to spatial ranges and times shorter than the size of the total

Laboratoire de Géophysique Interne et Tectonophysique, CNRS, Université de Savoie, 73376 Le Bourget du Lac, France.

*To whom correspondence should be addressed. E-mail: david.marsan@univ-savoie.fr

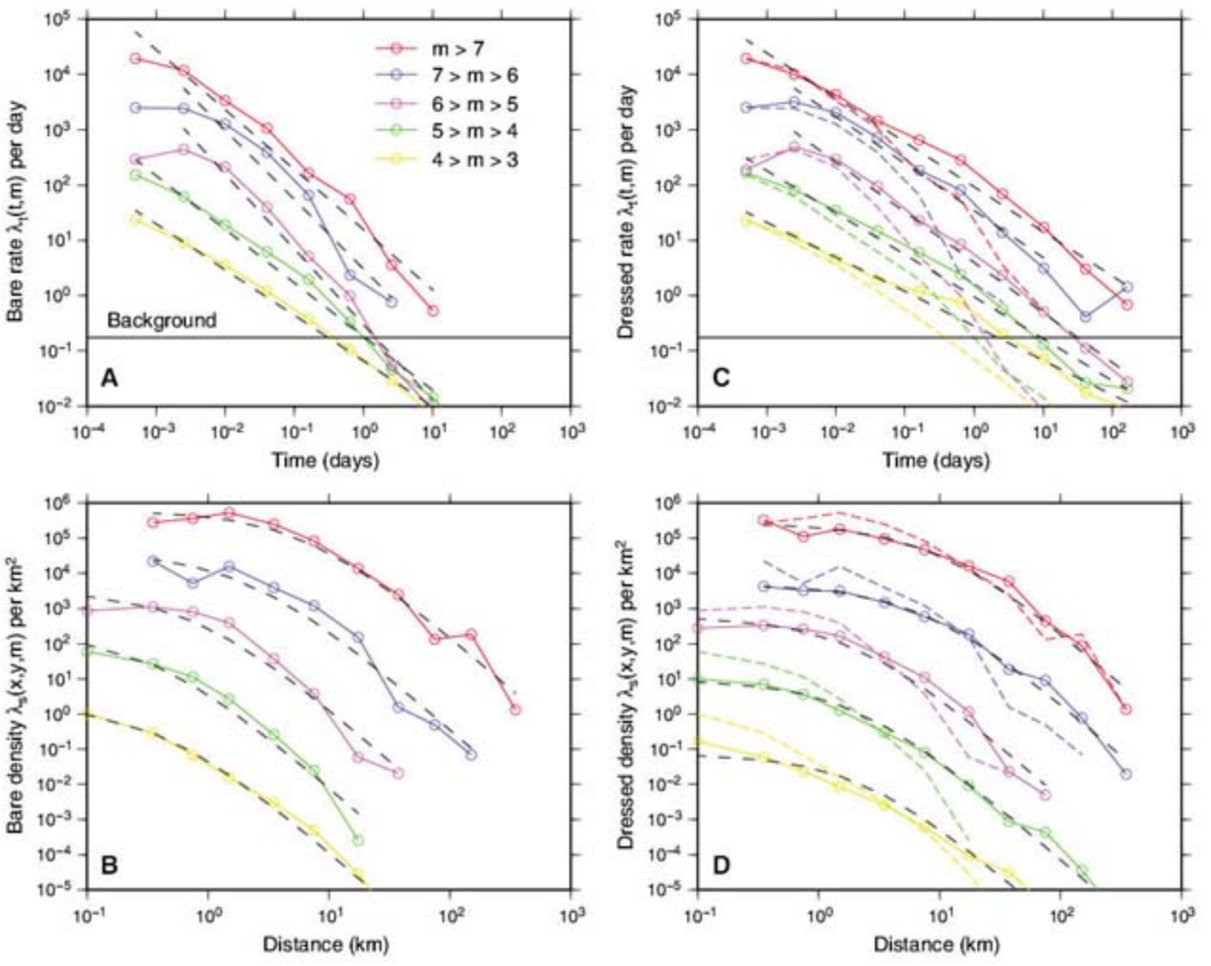


Fig. 1. Estimated rates and densities for California. (**A** and **B**) Bare kernels; (**C** and **D**) dressed kernels. The best power laws for the temporal rates $\lambda_t(t, m)$ and the best $[1 + (r/L)]^{-3}$ laws for the densities $\lambda_s(x, y, m)$ are shown as black dashed lines. The background temporal rate $\lambda_{0,t}$ [black horizontal line in (A) and (C)] is computed as $\sum_{i=1}^{N} w_{0,i}/T$. In (C) and (D), the dressed kernels (continuous lines) are compared to the bare ones (color dashed lines). The densities λ_s have been vertically shifted for clarity.

aftershock sequence. How much shorter is still an open question, in the absence of any simple or standard way to efficiently discriminate direct and indirect triggering in the data.

Seismologists mostly resort to declustering algorithms to separate earthquakes between mainshocks and aftershocks (*4*–*7*). These methods have arbitrary rules and are heavily parameter-dependent. Recently, more sophisticated methods were proposed to perform stochastic declustering—that is, determining the probability that earthquake A triggered earthquake B (*8*). This overseeds the usual approach of binary linking of one aftershock to one single mainshock: An earthquake j is then influenced by all preceding earthquakes i, according to influence weights $w_{i,j}$. These methods, however, are model-dependent, as the influence of a trigger earthquake is constrained to follow a specific law, whose parameters must be inverted.

Here we show that the probability of directly and indirectly triggering aftershocks can be estimated with no a priori model. A rapidly converging algorithm with a small number of hypotheses (linearity, mean-field) can decipher the complex seismicity time series to reveal the underlying triggering influences exerted by earthquakes of all sizes. A notable result is that even large earthquakes causally trigger aftershocks only during a relatively short time span. However, they condition regional seismicity for a much longer time period, and over larger, time-increasing distances, through the local triggering caused by their aftershocks. This cascading effect, dominated by small shocks, thus appears to be a crucial component in earthquake interactions.

Seismicity is considered as a point process in time, space, and magnitude. The observed (dressed) seismicity rate density $\lambda(x, t)$, defined as the number of earthquakes per unit time and unit area at position x and time t, is modeled as

$$\lambda(x,t) = \lambda_0 + \sum_{t_i<t} \lambda_i(x,t) \qquad (1)$$

where λ_0 is the uniform background rate density, and $\lambda_i(x, t)$ is the (bare) contribution of earthquake i that occurred at $\{x_i, t_i\}$, representing the aftershocks directly caused by this earthquake. We assume only that (i) the triggering process is linear [i.e., the bare contributions $\lambda_i(x, t)$ sum up], and (ii) a mean-field response to the occurrence of an earthquake can be estimated that depends only on its magnitude, $\lambda_i(x, t) = \lambda(|x - x_i|, t - t_i, m_i)$, hence two earthquakes of equal magnitude are modeled similarly.

The algorithm works as follows:

1. Knowing an a priori bare kernel $\lambda(|\Delta x|, \Delta t, m)$ and λ_0, we compute the triggering weights $w_{i,j} = \alpha_j \lambda(|x_j - x_i|, t_j - t_i, m_i)$ if $t_i < t_j$, $w_{i,j} = 0$ otherwise, and the background weights $w_{0,j} = \alpha_j \lambda_0$. The normalization coefficients α_j are such that

$$\sum_{i=1}^{j-1} w_{i,j} + w_{0,j} = 1 \qquad (2)$$

2. The updated bare rates are then computed as

$$\lambda(|\Delta x|, \Delta t, m) = \frac{1}{N_m \times \delta t \times S(|\Delta x|, \delta r)} \sum_{i,j \in A} w_{i,j} \qquad (3)$$

where A is the set of pairs such that $|x_i - x_j| = |\Delta x| \pm \delta r$, $m_i = m \pm \delta m$, and $t_j - t_i = t \pm \delta t$ (δr, δt, and δm are discretization parameters), N_m is the number of earthquakes such that $m_i = m \pm \delta m$, and $S(|\Delta x|, \delta r)$ is the surface covered by the disk with radii $|\Delta x| \pm \delta r$. The a posteriori background rate is

$$\lambda_0 = \frac{1}{T \times S} \sum_{j=1}^{N} w_{0,j} \qquad (4)$$

where T is the duration of the time series (containing N earthquakes) and S is the surface analyzed. This corresponds to stacking all the aftershocks following mainshocks i of similar magnitudes, but counting an aftershock j according to its weight $w_{i,j}$.

Starting with an initial guess for $\lambda(|\Delta x|, \Delta t, m)$, these two steps are iterated until

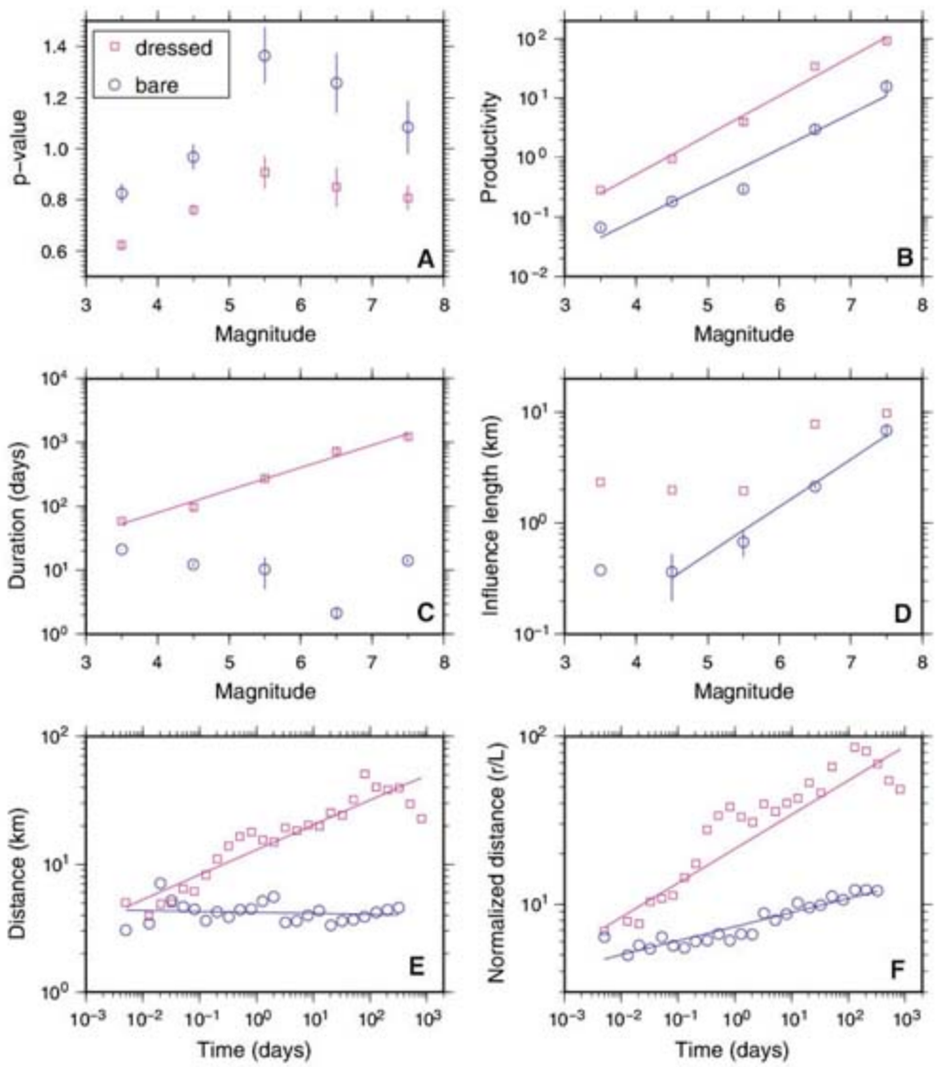


Fig. 2. Characteristics of bare (blue) and dressed (purple) aftershock sequences. (**A** and **B**) p values and productivity parameters χ from the fits $\lambda_t(t, m) = \chi t^{-p}$ of Fig. 1. Productivity grows as $10^{0.60 \pm 0.07m}$ (bare) and $10^{0.66 \pm 0.04m}$ (dressed). (**C**) Durations of the sequences. The dressed duration follows $10^{0.35m}$. (**D**) Influence lengths L estimated from the $[1 + (r/L)]^{-3}$ fits of the spatial densities. (**E**) Mean epicentral distance between mainshocks and aftershocks versus time following the mainshock. (**F**) Same as (E), for epicentral distance normalized by the bare influence length L of the mainshock. The best power laws $r \sim \delta t^H$ give (E) $H = -0.01 \pm 0.03$ (bare), H = 0.19 ± 0.04 (dressed), and (F) $H = 0.08 \pm 0.02$ (bare), H = 0.21 ± 0.06 (dressed).

convergence is reached—that is, until the weights (or the rates) do not change substantially during an iteration. Tests were conducted on synthetic catalogs, showing the ability of the algorithm to correctly estimate the cascade structure (*9*).

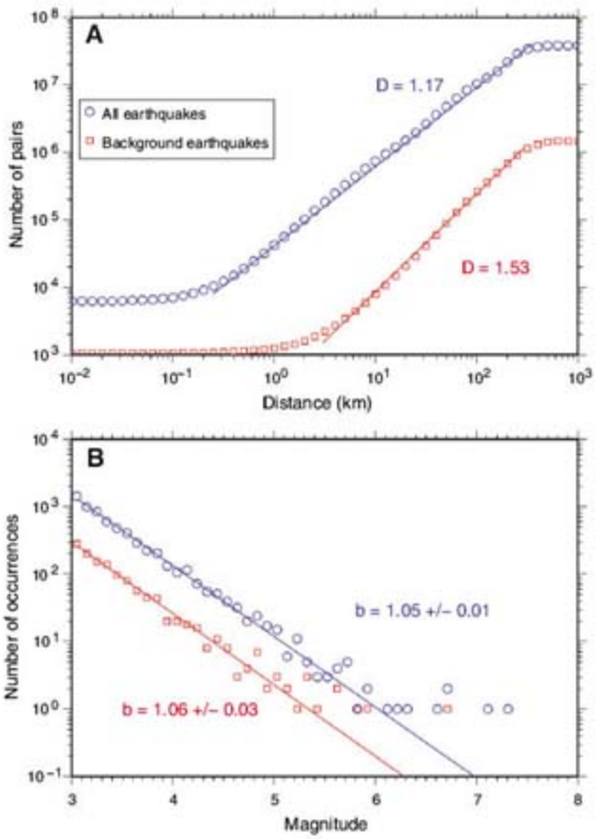


Fig. 3. Comparison of total (blue) and background (red) seismicity. (**A**) Number of pairs $N(r)$ with epicentral distance less than distance r, along with estimates of the fractal correlation dimension D such that $N(r) \sim r^D$. (**B**) Number of occurrences versus magnitude, scaling as $10^{-b \times m}$.

We analyzed seismicity in southern California (*10*) from 1 January 1984 to 31 December 2002. We considered only the $N = 6190$ $m \geq 3$ earthquakes in the catalog, both for completeness reasons and because the method uses a $N \times N$ weight matrix, preventing the analysis of too large a data set on a standard desktop computer. The completeness magnitude is estimated to be 2.2 for this catalog. The method was slightly modified to account for the fact that the background earthquakes are nonuniformly distributed in space (*9*).

The rates had roughly an Omori-Utsu decay $\lambda_t(t, m) = \chi t^{-p}$ (Fig. 1), where χ is the productivity. The p value increased with m (Fig. 2A), as observed previously (*11*) for dressed aftershock sequences, although with a saturation at $m \geq 6$. The rates decayed significantly more slowly when considering the full cascade including indirectly triggered aftershocks. The dressed p values were 0.2 to 0.4 units smaller than the bare p values (Fig. 2A). The productivity parameter grew as $= \chi \sim 10^{am}$, with $a = 0.60 \pm 0.07$ and $a = 0.66 \pm 0.04$ for the bare and dressed kernels, respectively (Fig. 2B). This yielded a significantly lower scaling exponent than previous estimates (*3*, *12*–*14*) for dressed sequences.

The densities were well fitted with a $\lambda_x(x, y, m) \sim [1 + (r/L)]^{-3}$ law, hence a r^{-2} decay of the number of aftershocks (Fig. 1). The bare

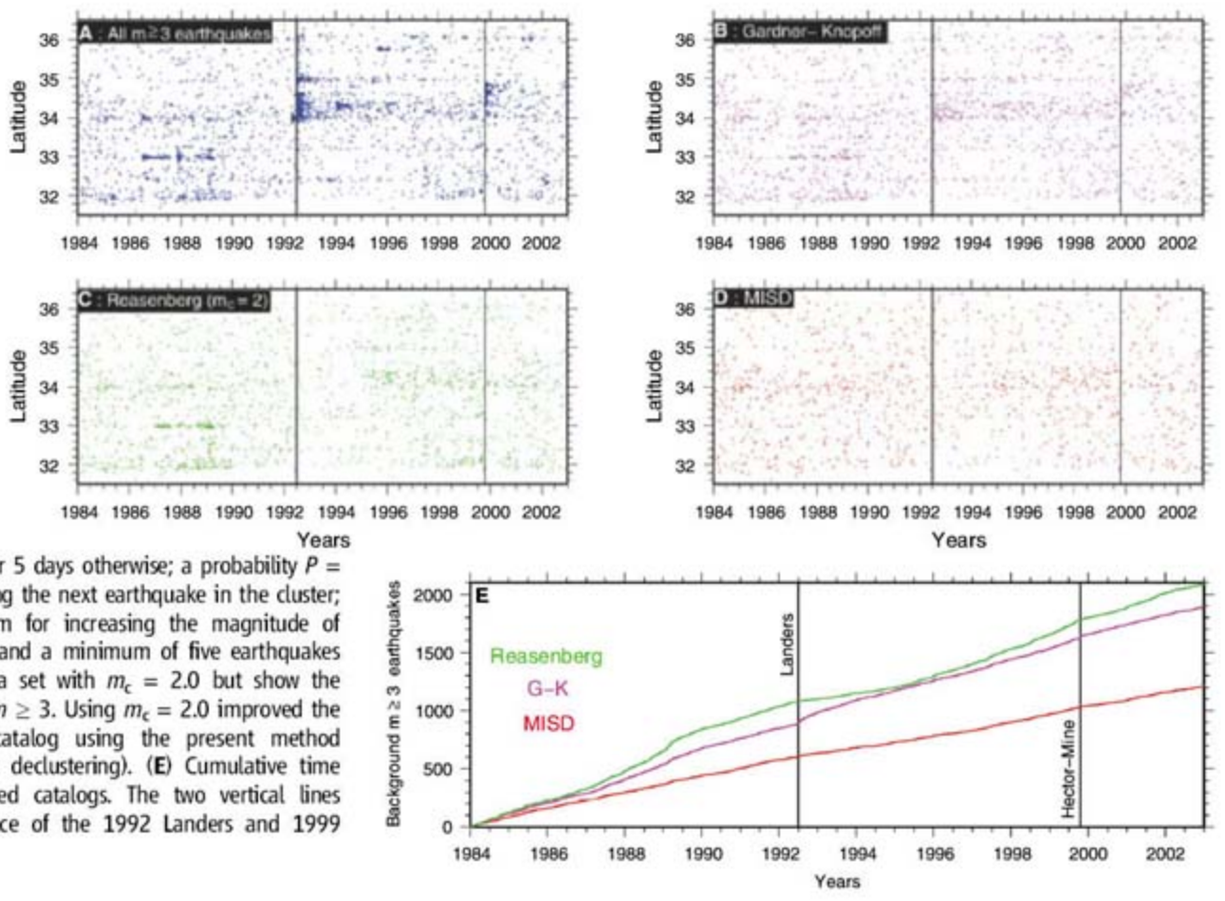


Fig. 4. Comparison of declustering algorithms. (**A** to **D**) Latitude versus time of occurrence of $m \geq 3$ earthquakes. (A) All earthquakes in the catalog. (B) Declustered catalog using the method of (*4*). (C) Declustered catalog using the algorithm of (*5*). Parameters are $r_{fact} = 8$, which gives the number of mainshock rupture lengths up to which a mainshock is supposed to trigger aftershocks; the maximum ahead-time for linking an aftershock to a cluster, which is 1 day if only one event belongs to the cluster, or 5 days otherwise; a probability $P = 95\%$ of confidence of observing the next earthquake in the cluster; $\chi_k = 0.5$, the correction term for increasing the magnitude of completeness during clusters; and a minimum of five earthquakes per cluster. We used the data set with $m_c = 2.0$ but show the declustered earthquakes with $m \geq 3$. Using $m_c = 2.0$ improved the algorithm. (D) Declustered catalog using the present method (model-independent stochastic declustering). (**E**) Cumulative time series of the three declustered catalogs. The two vertical lines indicate the time of occurrence of the 1992 Landers and 1999 Hector Mine earthquakes.

influence length L was remarkably small, ranging from 0.35 km at $4 \leq m < 5$ to 6.6 km at $m \geq 7$. It grew as $10^{0.43m}$ (Fig. 2D), which is close to the $10^{0.5m}$ dependence expected for the rupture length of small to intermediate-size earthquakes (*15*). The bare length was 0.35 km for $3 \leq m < 5$, which is due to the limited resolution on the relative hypocenter positions, as confirmed by the break in scaling at ~400 m for the correlation integral (Fig. 3A). The dressed influence lengths were about 5 times the bare ones. These influence lengths were not maximum triggering distances: Many aftershocks were triggered past L. Relaxing the point-like earthquake hypothesis affected these results for the large mainshocks. Using distances to the rupture plane rather than epicentral distances, the bare kernel of $m \geq 6$ mainshocks was moved toward longer-range triggering, greater aftershock productivity, and longer durations (*9*).

The durations of aftershock sequences (Fig. 2C) were computed by (i) calculating, for all the mainshocks, the delays after which the last direct and last indirect aftershocks occurred, and (ii) averaging these delays conditioned on the magnitude of the mainshock. The duration of direct aftershock sequence was largely independent of the mainshock magnitude and was generally short (on the order of 10 to 15 days for $m \geq 3$ aftershocks). The dressed sequence, however, lasted longer for larger mainshocks, following a $10^{0.35m}$ increase. This implies that short-lasting triggering mechanisms, acting at the time scale of a few days, could be the key process, along with the cascading effect, in controlling earthquake dynamics.

The slow expansion of aftershock zones has been reported in previous studies (*16*, *17*). We measured the mean distance r between mainshock and aftershock with time δt separating the two earthquakes. This distance was constant for bare aftershocks, whereas for dressed aftershocks it slowly grew as $r \sim \delta t^{0.19}$ (Fig. 2E). We reached similar conclusions when considering the distance normalized by the bare influence length L of the mainshock (Fig. 2F). This shows that cascading triggering drives the expansion of aftershock zones: The spatial pattern of direct triggering was almost constant with time, ruling out triggering by fluid movements or viscoelasticity at the time scales examined here.

The number of earthquakes directly triggered by all the earthquakes of a given magnitude slowly decreased with this magnitude, demonstrating the importance of small shocks in controlling the regional seismicity. The collective production scales as $10^{(a-b)m}$; here, $a = 0.6$ for the bare aftershocks (Fig. 2B) and $b = 1.05$ (Fig. 3B).

The background spatial function $g(r)$ decayed as $r^{-0.43}$, which is equivalent to saying that the background earthquake epicenters are fractally distributed with dimension $D = 1.57$. This is confirmed by the correlation integral of the background earthquakes (Fig. 3A), counting the number of pairs of earthquakes with distance less than a given value, each pair (i, j) being weighted by $w_{0,i} \times w_{0,j}$. The similar b values of the Gutenberg-Richter laws (*18*) for the total and the background earthquakes (Fig. 3B) suggest that the dynamic rupture extent is not a priori controlled by the triggering mechanism (either previous earthquakes or aseismic processes such as tectonic loading) at work.

We obtained a background rate of 0.17 $m \geq 3$ earthquakes per day in southern California over the years 1984 to 2002, which corresponds to 19.5% of the total rate of $m \geq 3$ earthquakes. The remaining ~80% can therefore be considered as resulting from stress transfer and fault interaction processes, causing the seismicity to be heavily clustered in time. Such a proportion at the magnitude cutoff $m_c = 3$ is coherent with the estimates in the range 18% to 24% we obtained using other nonparametric methods (*19*). The values computed using the present method, however, depend on m_c: Cutting at larger magnitudes increases the relative proportion of background earthquakes to 32% at $m_c = 4$ and 68% at $m_c = 5$. A larger cutoff causes the removal of small triggering earthquakes; earthquakes that were triggered by small shocks are then more likely to be seen as background earthquakes (*20*). Inversely, decreasing m_c would yield smaller percentages of background events. The 19.5% proportion at $m_c = 3$ is therefore an overestimation of the actual background contribution.

Declustering of earthquake catalogs aims at removing the aftershock clusters, keeping only statistically independent mainshocks. The declustering algorithms by Gardner and Knopoff (*4*) and Reasenberg (*5*) are the most classical methods. We ran these two methods along with ours (Fig. 4). The present method is better at identifying an underlying Poisson process, and it efficiently removes the aftershock clusters following large mainshocks such as the 1992 Landers and 1999 Hector Mine earthquakes. Moreover, it does not rely on any parameterization.

Cascading of aftershock triggering is an essential component of seismicity. It has a scale-invariant structure, making earthquake declustering an ill-defined problem: The statistical dependence between earthquakes increases when decreasing the value of m_c, so that the remaining set of declustered mainshocks heavily depends on m_c. Because of this cascading, the aftershock sequence initiated by a mainshock is substantially extended, mostly in time. Conversely, what appears at first as an aftershock cluster related to a well-identified mainshock is in fact mostly caused not by the mainshock itself, but rather by intermediate aftershocks. When decreasing m_c, the direct triggering effect due to large mainshocks could potentially be even further reduced relative to direct triggering by small shocks. This is particularly critical for understanding the physical mechanisms that cause earthquake triggering: The testing and validation of models first require the correct relation of the aftershocks to their trigger, rather than to an older, generally bigger, ancestor in the triggering chain.

References and Notes

1. K. R. Felzer, T. W. Becker, R. E. Abercrombie, G. Ekström, J. R. Rice, *J. Geophys. Res.* **107**, 2190 (2002).
2. D. Marsan, *Geophys. J. Int.* **163**, 141 (2005).
3. A. Helmstetter, Y. Y. Kagan, D. D. Jackson, *J. Geophys. Res.* **110**, B05S08 (2005).
4. J. K. Gardner, L. Knopoff, *Bull. Seismol. Soc. Am.* **64**, 1363 (1974).
5. P. Reasenberg, *J. Geophys. Res.* **90**, 5479 (1985).
6. S. D. Davis, C. Frohlich, *Geophys. J. Int.* **104**, 289 (1991).
7. G. M. Molchan, O. E. Dmitrieva, *Geophys. J. Int.* **109**, 501 (1992).
8. J. Zhuang, Y. Ogata, D. Vere-Jones, *J. Am. Stat. Assoc.* **97**, 369 (2002).
9. See supporting material on *Science* Online.
10. P. Shearer, E. Hauksson, G. Lin, *Bull. Seismol. Soc. Am.* **95**, 904 (2005).
11. G. Ouillon, D. Sornette, *J. Geophys. Res.* **110**, B04306 (2005).
12. B. E. Shaw, *Geophys. Res. Lett.* **20**, 907 (1993).
13. A. Helmstetter, *Phys. Rev. Lett.* **91**, 058501 (2003).
14. K. R. Felzer, R. E. Abercrombie, G. Ekström, *Bull. Seismol. Soc. Am.* **94**, 88 (2004).
15. C. H. Scholz, *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, Cambridge, 1991).
16. F. Tajima, H. Kanamori, *Phys. Earth Planet Inter.* **40**, 77 (1985).
17. A. Helmstetter, G. Ouillon, D. Sornette, *J. Geophys. Res.* **108**, 2483 (2003).
18. B. Gutenberg, C. F. Richter, *Bull. Seismol. Soc. Am.* **142**, 185 (1944).
19. S. Hainzl, F. Scherbaum, C. Beauval, *Bull. Seismol. Soc. Am.* **96**, 313 (2006).
20. D. Sornette, M. J. Werner, *J. Geophys. Res.* **110**, B09303 (2005).

21. We thank G. Daniel, A. Helmstetter, G. Ouillon, and D. Sornette for constructive discussions. Supported by European Community project VOLUME (FP6-2004-Global-3) (O.L.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1076/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
Table S1
References

3 August 2007; accepted 10 January 2008
10.1126/science.1148783

Human-Induced Changes in the Hydrology of the Western United States

Tim P. Barnett,[1]* David W. Pierce,[1] Hugo G. Hidalgo,[1] Celine Bonfils,[2] Benjamin D. Santer,[2] Tapash Das,[1] Govindasamy Bala,[2] Andrew W. Wood,[3] Toru Nozawa,[4] Arthur A. Mirin,[2] Daniel R. Cayan,[1,5] Michael D. Dettinger[1,5]

Observations have shown that the hydrological cycle of the western United States changed significantly over the last half of the 20th century. We present a regional, multivariable climate change detection and attribution study, using a high-resolution hydrologic model forced by global climate models, focusing on the changes that have already affected this primarily arid region with a large and growing population. The results show that up to 60% of the climate-related trends of river flow, winter air temperature, and snow pack between 1950 and 1999 are human-induced. These results are robust to perturbation of study variates and methods. They portend, in conjunction with previous work, a coming crisis in water supply for the western United States.

Water is perhaps the most precious natural commodity in the western United States. Numerous studies indicate the hydrology of this region is changing in ways that will have a negative impact on the region (*1*–*3*). Between 1950 and 1999 there was a shift in the character of mountain precipitation, with more winter precipitation falling as rain instead of snow (*2*, *4*, *5*), earlier snow melt (*4*, *6*), and associated changes in river flow (*7*–*10*). In the latter case, the river flow experiences relative increases in the spring and relative decreases in the summer months. These effects go along with a warming over most of the region that has exacerbated these drier summer conditions (*5*, *8*, *11*).

The west naturally undergoes multidecadal fluctuations between wet and dry periods (*12*). If drying from natural climate variability is the cause of the current changes, a subsequent wet period will likely restore the hydrological cycle to its former state. But global and regional climate models forced by anthropogenic pollutants suggest that human influences could have caused the shifts in hydrology (*2*, *13*–*15*). If so, these changes are highly likely to accelerate, making modifications to the water infrastructure of the western United States a virtual necessity.

Here, we demonstrate statistically that the majority of the observed low-frequency changes in the hydrological cycle (river flow, temperature, and snow pack) over the western United States from 1950 to 1999 are due to human-caused climate changes from greenhouse gases and aerosols. This result is obtained by evaluating a combination of global climate and regional hydrologic models, together with sophisticated data analysis. We use a multivariable detection and attribution (D&A) methodology (*16*–*18*) to show that the simultaneous hydroclimatic changes observed already differ significantly in length and strength from trends expected as a result of natural variability (detection) and differ in the specific ways expected of human-induced effects (attribution). Focusing on the hydrological cycle allows us to assess the origins of the most relevant climate change impacts in this water-limited region.

We investigated simultaneous changes from 1950 to 1999 (*19*) in snow pack (snow water equivalent or SWE), the timing of runoff of the major western rivers, and average January through March daily minimum temperature (JFM T_{min}) in the mountainous regions of the western United States (*20*). These three variates arguably are among the most important metrics of the western hydrological cycle. By using the multivariable approach, we obtain a greater signal-to-noise (S/N) ratio than from univariate D&A alone (see below).

The SWE data are normalized by October-to-March precipitation (P) to reduce variability from heavy- or light-precipitation years. Observed SWE/P and temperature were averaged over each of nine western mountainous regions (Fig. 1) to reduce small-spatial-scale weather noise. The river flow variate is the center of timing (CT), the day of the year on which one-half of the total water flow for the year has occurred, computed from naturalized flow in the Columbia, Colorado, and Sacramento/San Joaquin rivers. CT tends to decrease with warming because of earlier spring melting.

Selected observations from these regions and variables are displayed in Fig. 2, showing the trends noted above, along with substantial regional differences and "weather noise." SWE/P trends in the nine regions vary from –2.4 to –7.9% per decade, except in the southern Sierra Nevada where the trend is slightly positive. The JFM T_{min} trends are all positive and range from 0.28° to 0.43°C per decade, whereas the river CT

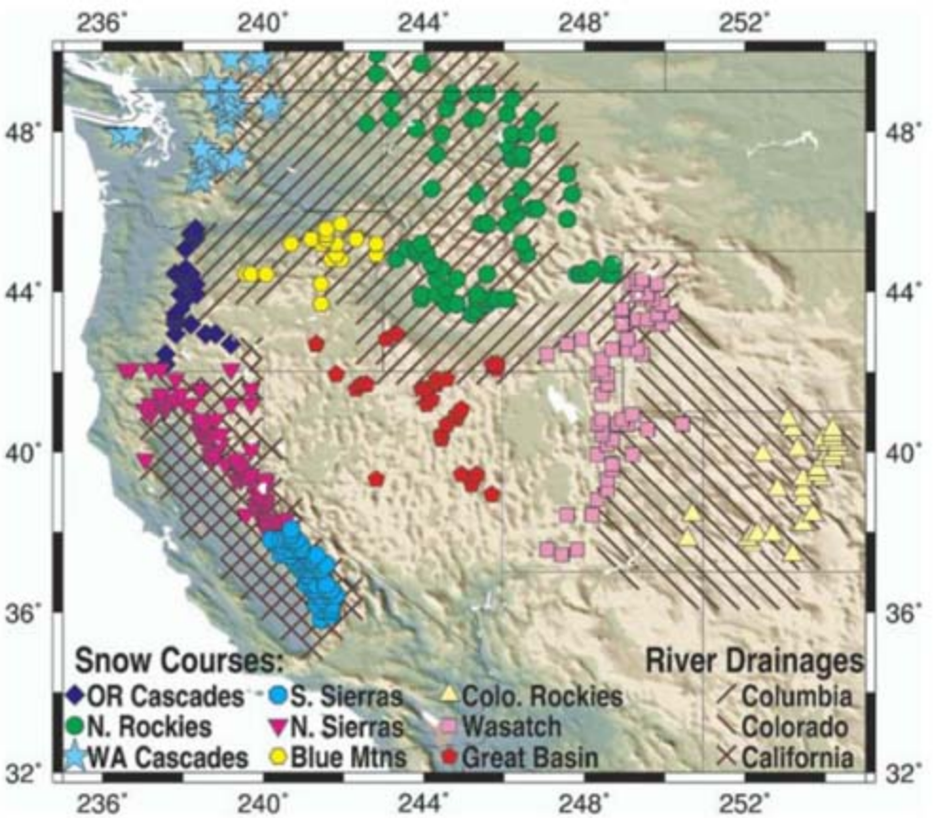


Fig. 1. Map showing averaging regions over which SWE/P and JFM T_{min} were determined. The hatching shows the approximate outline of the three main drainage basins used in this study.

[1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA. [2]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [3]Land Surface Hydrology Research Group, Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA. [4]National Institute for Environmental Studies, 16-2, Onogawa, Tsukuba, Ibaraki 305-8506, Japan. [5]U.S. Geological Survey, La Jolla, CA 92093, USA.

*To whom correspondence should be addressed. E-mail: tbarnett-ul@ucsd.edu

arrives from 0.3 to 1.7 days per decade earlier. The challenge in D&A analysis is to determine whether a specific, predetermined signal representing the response to external forcing is present in these observations.

We compared the observations with results from a regional hydrologic model forced by global climate model runs. One of the global models, the Parallel Climate Model (PCM) (*21*), has been used previously in hydrological studies in the western United States (*22*) and realistically portrays important features of observed climate and the amplitude of natural internal variability. The second climate model, the anthropogenically forced medium-resolution Model for Interdisciplinary Research on Climate (MIROC) (*23*–*25*), was selected from the current Intergovernmental Panel on Climate Change (IPCC) AR4 set of global runs (*26*) because it had available many 20th-century ensemble members with daily data, and because it offered a high degree of realism in representing the Pacific Decadal Oscillation (PDO). We used the anthropogenically forced versions of these models to obtain an estimate for the expected signal not confounded by other forcing mechanisms. The models provided multiple realizations (10 for MIROC, 4 for PCM) of the historical response of the climate system to anthropogenic forcing. The daily output from these coarse-horizontal-resolution model results was downscaled to a 1/8° × 1/8° latitude-longitude grid by two different statistical methods [Bias Correction and Spatial Disaggregation (BCSD) (*27*) and Constructed Analogues (CA) (*28*)]. The downscaled temperature and precipitation data were supplied as input to the Variable Infiltration Capacity (VIC) hydrological model (*15*, *27*, *29*) to obtain river flow and SWE/P.

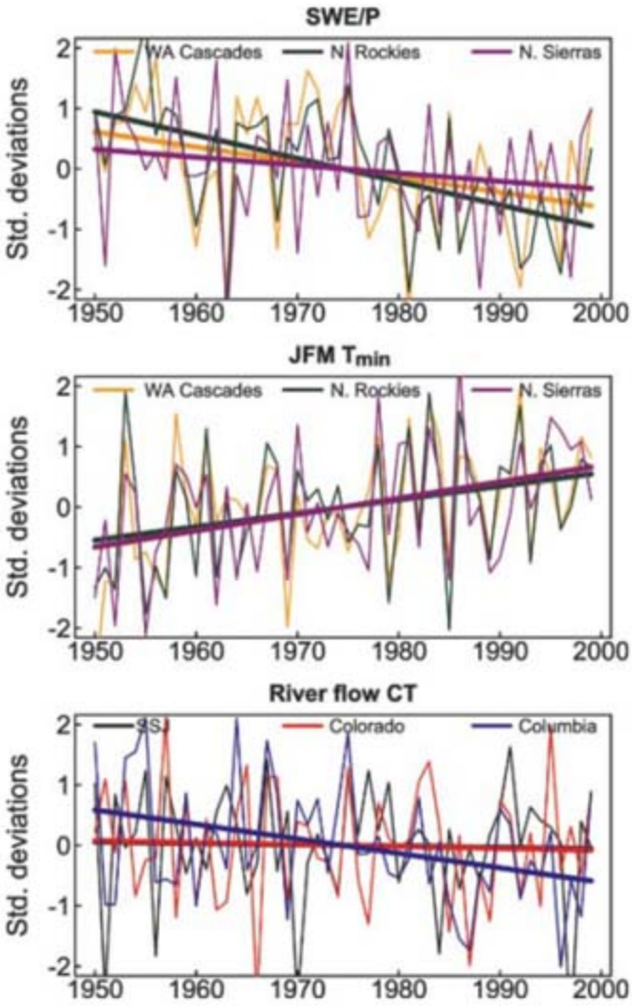


Fig. 2. Observed time series of selected variables (expressed as unit normal deviates) used in the multivariate detection and attribution analysis. Taken in isolation, seven of nine SWE/P, seven of nine JFM T_{min}, and one of the three river flow variables have statistically significant trends.

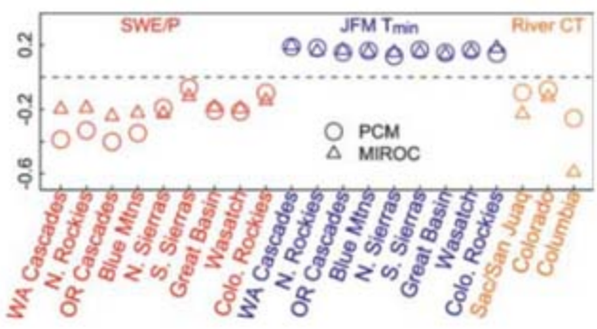


Fig. 3. Fingerprints from the multivariate analysis of PCM and MIROC.

We used the downscaled model results to estimate an anthropogenic "fingerprint" for the PCM and MIROC models (*30*). The fingerprint describes the joint variability of SWE/P, JFM T_{min}, and river flow (Fig. 3) (*20*). The model fingerprints are very similar despite the different external forcings used (*20*, *26*). The results show that warmer temperatures accompany decreases in SWE/P and decreases in CT of major western river systems. The sign of each variable is a monopole, indicating a coherent regional-scale signal over the western United States.

The temporal component of the fingerprint (not shown) is well represented by a simple trend. This implies that the fingerprint primarily captures the spatial expression of long-term changes, and not shorter-period climate modes (such as El Niño–Southern Oscillation or the PDO).

The signal strength is calculated as the least-squares linear trend of the projection of a data set (model or observations) onto the fingerprint (*20*). The upper panel of Fig. 4 shows the ensemble mean signals for our various model runs and the observations (*20*). The observations show a positive signal indistinguishable from the PCM and MIROC anthropogenically forced runs. These signals exclude zero at the 95% confidence interval, thus achieving "detection."

We used 1600 years of downscaled control run data from two different global models (*20*) to estimate the probability that the observed signal could be due to natural, internal variability (Fig. 4, lower panel). The observed signal falls outside the range expected from natural variability with high confidence ($P < 0.01$). In separate analyses for PCM and MIROC, the likelihood that the model signal arises from natural internal variability is between 0.01 and 0.001 (*20*). The different downscaling methods have little impact on these results. We conclude that natural internal climate variability alone cannot explain either the observed or simulated changes in SWE/P, JFM T_{min}, and CT in response to anthropogenic forcing.

PCM simulations forced solely by the combined impacts of observed solar variability and volcanic activity (Sol/Vol, Fig. 4) show a signal with sign opposite to that observed. We conclude that solar and volcanic forcing also fail to explain the observed hydrological changes.

Might anthropogenically induced precipitation changes account for our results? This is unlikely because our variables were chosen to minimize sensitivity to precipitation fluctuations. However, previous work has identified an anthropogenic effect on global-scale changes in precipitation (*31*). We conducted a univariate D&A analysis on precipitation, comparing the fingerprint obtained from the anthropogenic runs to the control runs and observations. The results (Fig. 4, lower panel) show that the observed changes in precipitation over the nine western U.S. mountain regions are indistinguishable from natural variability. We found the same for model precipitation (not shown). We conclude that although

precipitation may be affected by anthropogenic forcing on larger scales or in other regions, or in this region in the future, it cannot explain the strong changes in western U.S. hydrology from 1950 to 1999.

Finally, the observations are consistent with the anthropogenic model runs. The observed signal is stronger than found in either model, but the differences are not statistically significant. The ensemble mean signal strength from PCM is 60% of the observed signal strength; that is, PCM estimates that three-fifths of the projected trend can be ascribed to human effects. The two downscaling methods give somewhat different signal strengths (Fig. 4), but the attribution holds no matter which is chosen. We conclude that application of a rigorous, multivariable D&A methodology shows a detectable and attributable signature of human effects on western hydrology.

We examined the time evolution of signal and noise by projecting the observations (signal) and control run data (noise) onto the multivariable fingerprint, then fitting linear trends of increasing length L to the resulting projected time series. This enabled us to calculate a S/N ratio as a function of L (from 10 to 50 years); Fig. 5 shows that the S/N ratio rises above the 5% significance threshold no later than 1986. This result is robust to uncertainties in the model fingerprint, model-based noise estimates, and statistical downscaling method (*20*). We also repeated the D&A analysis without areal weighting and found that it made no difference in our conclusions.

The variables examined here covary in a physically and internally consistent way: An increase in minimum temperature is associated with less SWE/P and earlier runoff. Quantitatively, we also compared the S/N obtained from separate analyses of each variable with that obtained for the full multivariable problem (*20*). For fixed choices of fingerprint, noise, and downscaling (*32*), the S/N values from the separate SWE/P, JFM T_{min}, and CT analyses were 2.90, 2.95, and 1.85, respectively, all significant at about the 0.05 level or above. The multivariable analysis had a S/N of 3.62, and so it has quantitative value as well as providing a test of whether SWE/P, JFM T_{min}, and CT covary in a physically consistent way.

Our results are robust with respect to uncertainties in model estimates of anthropogenic climate fingerprints and natural variability, downscaling method, and the choice of univariate or multivariate D&A analysis. Estimates of natural variability used for significance testing agree well with those derived from paleo proxies (*20*). The analyses show with high confidence that the majority of the detrimental changes already seen in western U.S. hydrology are caused by human-induced effects. PCM, which has the most realistic signal strength, shows that human effects account for 60% of the observed 1950–1999 trend in signal strength. MIROC accounts for 35% of the trend. On the basis of Fig. 4 (upper panel) and the discussion of MIROC in (*20*), the PCM number seems more reliable.

Our results are not good news for those living in the western United States. The scenario for how western hydrology will continue to change has already been published using one of the models used here [PCM (*2*)] as well as in other recent studies of western U.S. hydrology [e.g., (*15*)]. It foretells water shortages, lack of storage capability to meet seasonally changing river flow, transfers of water from agriculture to urban uses, and other critical impacts. Because PCM performs so well in replicating the complex signals of the last half of the 20th century, we have every reason to believe its projections and to act on them in the immediate future.

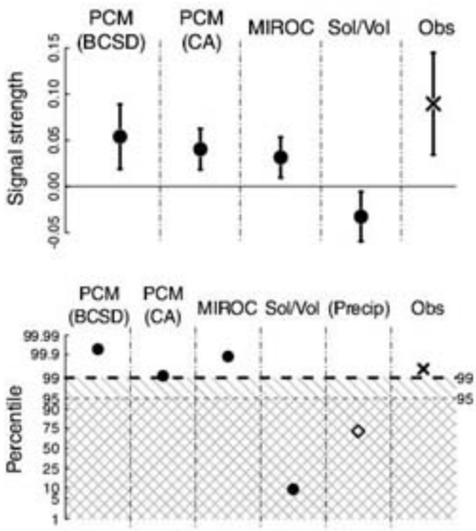


Fig. 4. Ensemble average signal strength (upper panel; standard deviations of the fingerprint's principal component per decade) and percentile rank of ensemble mean signal strength for the indicated model runs with respect to the combined (CCSM3-FV and PCM) control run (lower panel). Percentile values were calculated by Monte Carlo resampling of the control run taking into account N, the varying number of ensemble members. PCM (BCSD) and PCM (CA): PCM runs with anthropogenic forcing, with two different downscaling methods as described in the text (N = 4). MIROC: MIROC runs with anthropogenic forcing (N = 10). Sol/Vol: PCM runs with only solar and volcanic forcing included (N = 2). The cross shows the signal strength obtained from the observations (N = 1). For comparison purposes, also shown is the observed signal strength from a separate analysis of precipitation changes over the nine mountain regions (diamond). Values outside the hatched and cross-hatched regions are significant at the 0.01 and 0.05 levels, respectively.

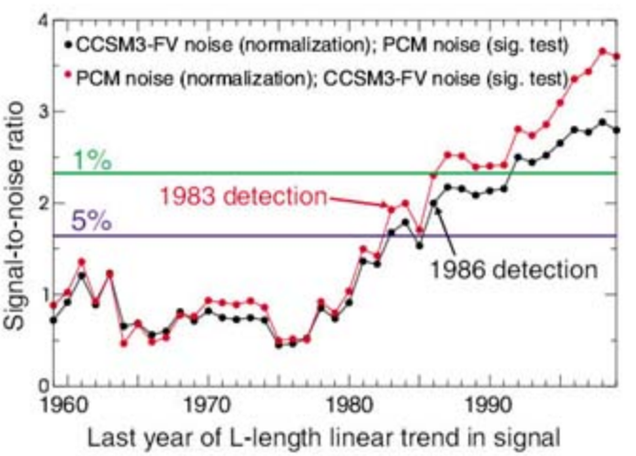


Fig. 5. Time-dependent S/N estimates for two different estimates of natural variability. The x axis is the last year of L-length linear trend in the signal estimate.

References and Notes

1. P. Gleick, *Water Resour. Res.* **23**, 1049 (1987).
2. ACPI, The Accelerated Climate Prediction Initiative, *Clim. Change* **62**, 444 (2004).
3. B. Udall, G. Bates, *Intermountain West Climate Summary, Western Water Assessment*, January 2007 (available from University of Colorado).
4. A. F. Hamlet, P. W. Mote, M. P. Clark, D. P. Lettenmaier, *J. Clim.* **18**, 4545 (2005).
5. N. Knowles, M. D. Dettinger, D. R. Cayan, *J. Clim.* **19**, 4545 (2006).
6. P. W. Mote, A. F. Hamlet, M. P. Clark, D. P. Lettenmaier, *Bull. Am. Meteorol. Soc.* **86**, 39 (2005).
7. M. D. Dettinger, D. R. Cayan, *J. Clim.* **8**, 606 (1995).
8. D. R. Cayan, S. Kammerdiener, M. D. Dettinger, J. Caprio, D. Peterson, *Bull. Am. Meteorol. Soc.* **82**, 399 (2001).
9. I. T. Stewart, D. R. Cayan, M. D. Dettinger, *J. Clim.* **18**, 1136 (2005).
10. S. K. Regonda, B. Rajagopalan, M. Clark, J. Pitlick, *J. Clim.* **18**, 372 (2005).
11. P. Y. Groisman *et al.*, *J. Hydrometeorol.* **5**, 64 (2003).
12. Colorado River Basin Water Management, *Evaluating and Adjusting to Hydroclimatic Variability* (National Academy of Sciences, Washington, DC, 2007).
13. P. C. D. Milly, K. A. Dunne, A. V. Vecchia, *Nature* **438**, 347 (2005).
14. R. Seager *et al.*, *Science* **316**, 1181 (2007); published online 4 April 2007 (10.1126/science.1139601).
15. N. Christiansen, D. Lettenmaier, *Hydrol. Earth Syst. Sci. Discuss.* **3**, 1 (2006).
16. T. P. Barnett, M. Schlesinger, *J. Geophys. Res.* **92**, 14772 (1987).
17. B. D. Santer *et al.*, *Clim. Dyn.* **12**, 77 (1995).
18. R. Schnur, K. I. Hasselmann, *Clim. Dyn.* **24**, 45 (2005).
19. Note that this period excludes the large-scale changes in runoff, precipitation, and water storage that have occurred in the southwest, especially the Colorado River drainage, since 2000. We do not claim that the large changes since 2000 are necessarily the result of human-induced warming.
20. See supporting material on *Science* Online.
21. W. Washington *et al.*, *Clim. Dyn.* **16**, 755 (2000).
22. T. P. Barnett *et al.*, *Clim. Change* **62**, 1 (2004).
23. K-1 Model Developers, *K-1 Coupled Model (MIROC) Description (K-1 Technical Report 1)*, H. Hasumi, S. Emori, Eds. (Center for Climate System Research, University of Tokyo, 2004).
24. T. Nozawa *et al.*, *Geophys. Res. Lett.* **32**, L20719 (2005).
25. T. Nozawa *et al.*, *MIROC, CGER's Supercomputer Monograph Report 12* (Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan, 2007).

26. B. D. Santer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15248 (2007).
27. A. Wood *et al.*, *Clim. Dyn.* **16**, 755 (2004).
28. E. P. Maurer, H. G. Hidalgo, *Hydrol. Earth Syst. Sci. Discuss.* **4**, 3413 (2007).
29. X. Liang, D. Lettermaier, A. Wood, S. Burges, *J. Geophys. Res.* **99**, 14415 (1994).
30. T. P. Barnett *et al.*, *Science* **309**, 284 (2005); published online 2 June 2005 (10.1126/science.1112418).
31. X. Zhang *et al.*, *Nature* **448**, 461 (2007).
32. The choices were CCSM3-FV noise for significance testing, PCM fingerprint, and statistical downscaling with the CA method. In the multivariable case, PCM noise was used for normalization.

33. The MIROC data were generously supplied by the National Institute for Environmental Studies, Onogawa, Tsukuba, Ibaraki, Japan. The PCM simulation had previously been made available to the Scripps Institution of Oceanography (SIO) by the National Center for Atmospheric Research for the Accelerated Climate Prediction Initiative project. Supported by the Lawrence Livermore National Laboratory (LLNL) through a Laboratory-Directed Research and Development grant to SIO via the San Diego Super Computer Center for the LUCSiD project; the U.S. Department of Energy and NOAA through the International Detection and Attribution Group (T.P.B.); Program of Climate Model Diagnoses and Intercomparison grant DOE-W-7405-ENG-48 (C.B., B.D.S., G.B., A.A.M.); the U.S. Geological Survey and SIO (D.R.C., M.D.D.); and the California Energy Commission (D.W.P., H.G.H.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1152538/DC1
SOM Text
Figs. S1 to S3
References

2 November 2007; accepted 23 January 2008
Published online 31 January 2008;
10.1126/science.1152538
Include this information when citing this paper.

Atomic-Level Models of the Bacterial Carboxysome Shell

Shiho Tanaka,[1] Cheryl A. Kerfeld,[3,4] Michael R. Sawaya,[2] Fei Cai,[5] Sabine Heinhorst,[5] Gordon C. Cannon,[5] Todd O. Yeates[1,2]*

The carboxysome is a bacterial microcompartment that functions as a simple organelle by sequestering enzymes involved in carbon fixation. The carboxysome shell is roughly 800 to 1400 angstroms in diameter and is assembled from several thousand protein subunits. Previous studies have revealed the three-dimensional structures of hexameric carboxysome shell proteins, which self-assemble into molecular layers that most likely constitute the facets of the polyhedral shell. Here, we report the three-dimensional structures of two proteins of previously unknown function, CcmL and OrfA (or CsoS4A), from the two known classes of carboxysomes, at resolutions of 2.4 and 2.15 angstroms. Both proteins assemble to form pentameric structures whose size and shape are compatible with formation of vertices in an icosahedral shell. Combining these pentamers with the hexamers previously elucidated gives two plausible, preliminary atomic models for the carboxysome shell.

The carboxysome enhances CO_2 fixation inside many photosynthetic and chemoautotrophic bacterial cells by encapsulating the key enzymes ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) and carbonic anhydrase (*1*–*3*). In contrast to membrane-bound eukaryotic organelles, carboxysomes and related bacterial microcompartments (*4*, *5*) have a proteinaceous outer shell, which is roughly polyhedral in shape (Fig. 1). The earliest observations of carboxysomes by electron microscopy date back more than 40 years (*6*, *7*). Subsequent genetic and biochemical studies have provided essential information about their function and composition (*3*, *8*–*10*); structural studies (*11*–*14*) have begun to illuminate the functional mechanisms and architectural details of the carboxysome.

Carboxysomes are found in all cyanobacteria and in some chemoautotrophic bacteria. Two types of carboxysomes have been defined by patterns of genomic organization and by protein sequence comparisons (*15*). They are represented in two model organisms, *Halothiobacillus neapolitanus*

[1]Department of Chemistry and Biochemistry, University of California at Los Angeles (UCLA), Los Angeles, CA 90095, USA. [2]UCLA–Department of Energy Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA. [3]Department of Energy–Joint Genome Institute, Walnut Creek, CA 94598, USA. [4]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. [5]Department of Chemistry and Biochemistry, University of Southern Mississippi, Hattiesburg, MS 39406, USA.

*To whom correspondence should be addressed. E-mail: yeates@mbi.ucla.edu

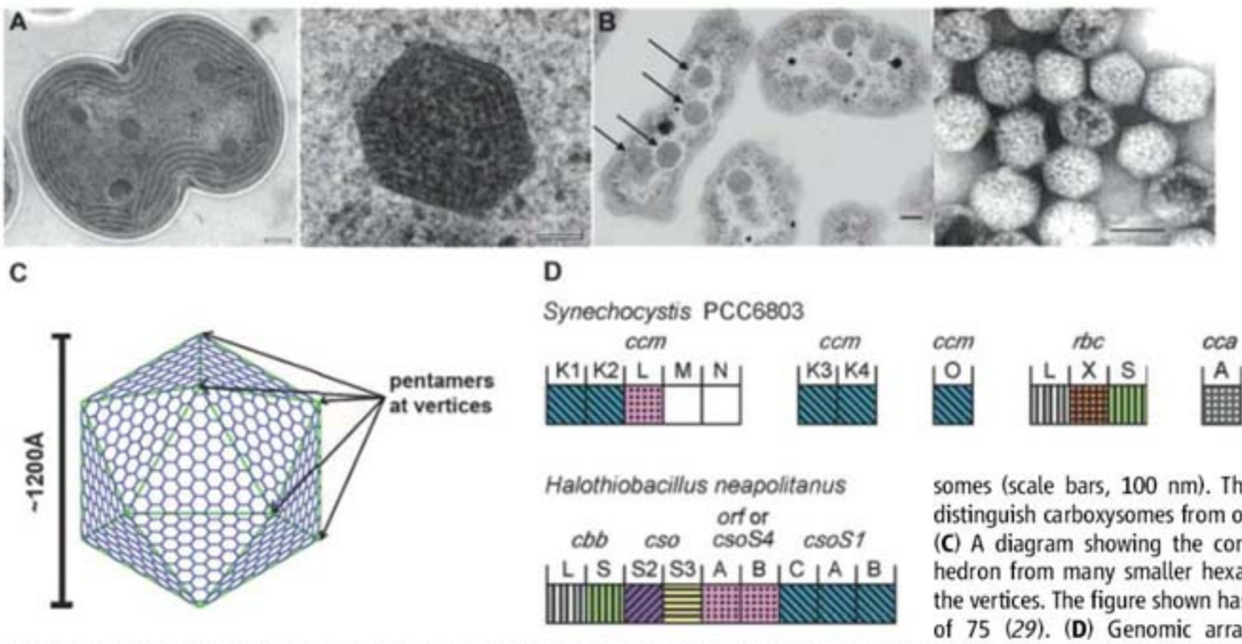


Fig. 1. Carboxysome architecture and operon organization. Transmission electron micrographs showing (**A**) (left) a section through a dividing cyanobacterial cell, *Syn.* 6803 (scale bar, 200 nm), and (right) an enlargement of a single carboxysome (scale bar, 50 nm) and (**B**) (left) *Halothiobacillus neapolitanus* cells with their carboxysomes highlighted by arrows and (right) purified carboxysomes (scale bars, 100 nm). Their polyhedral shape helps distinguish carboxysomes from other cytoplasmic inclusions. (**C**) A diagram showing the construction of a large icosahedron from many smaller hexagons and 12 pentagons at the vertices. The figure shown has a triangulation number (T) of 75 (*29*). (**D**) Genomic arrangement of carboxysome-associated proteins. Homologous proteins from the two model organisms are shaded and colored similarly. The hexameric shell proteins are colored in blue. The CcmL and OrfA (or CsoS4A) proteins under study here are colored in pink. The *rbc* and *cbb* genes code for the large and small subunit of RuBisCO. (A) is adapted from ref. (*11*); (B) is adapted from ref. (*12*).

(a chemoautotroph containing α-carboxysomes) and *Synechocystis* sp. strain PCC (Pasteur Culture Collection) 6803 (*Syn.* 6803, a cyanobacterium containing β-carboxysomes); potential functional differences between the two types are not clear. The carboxysome shell is composed predominantly of CsoS1 proteins (homologous proteins CsoS1A, B, and C) in *H. neapolitanus* (*16*, *17*) and of CcmK proteins (homologous proteins CcmK1 to 4) in *Syn.* 6803 (Fig. 1D). We have previously determined the structures of four of these shell proteins and found that they are all hexamers and that they tend to form extended, tightly packed molecular layers hypothesized to represent the flat facets of the polyhedral carboxysome shell. The structures suggest that diffusion of metabolites into and out of the carboxysome is limited by the small hexameric pores and gaps between the hexamers (*11*, *12*).

Two recent electron microscopy studies have confirmed that carboxysomes are approximately icosahedral in shape (*13*, *14*). The construction of large icosahedral structures typically requires a combination of hexameric and pentameric units (*18*, *19*). Pentamers generate curvature in an otherwise flat hexagonal sheet and occupy the vertices of an icosahedral shell (Fig. 1C). Candidates for such pentameric components of the shell had not been identified. Functions have been elucidated for most of the proteins encoded within well-characterized carboxysome operons (*3*). However, functions had not been assigned to the gene products of two open reading frames in *H. neapolitanus*, designated *orfA* and *orfB*, or their homolog in *Syn.* 6803, CcmL (Fig. 1D).

The CcmL protein from *Syn.* 6803 was expressed in *Escherichia coli* cells, and its crystal structure was determined by selenomethionine multiwavelength anomalous phasing methods (*20*). An atomic model was built and refined at a resolution of 2.4 Å, with a final model consisting of 96 of the 100 residues of the complete CcmL protein (*20*). OrfA was expressed in *E. coli* cells, and its crystal structure was determined by molecular replacement using the CcmL structure as a search model; the amino acid sequence identity between the two proteins is 37%. The OrfA structure was refined at a resolution of 2.15 Å, yielding a final model that contains 82 of the 83 amino acid residues in that protein (*20*). The structures of CcmL and OrfA are highly similar; their backbones superimpose with a root mean square deviation (RMSD) of only 1.0 Å (Fig. 2). CcmL consists of seven β strands and one α helix, with five of the β strands forming a β barrel (Fig. 2). OrfA contains the same central five-stranded β barrel and the α helix, but it lacks the two C-terminal β strands present in CcmL, owing to its smaller size.

Both CcmL and OrfA formed symmetric pentamers in their respective crystal structures (Fig. 2). The individual subunits of the pentamers are held tightly together by their protruding C-terminal regions, which form extensive interactions with neighboring subunits. The large subunit interfaces and the structural agreement between the two proteins argue that the observed pentameric structures represent their natural biological forms. The CcmL pentamer is shaped roughly like a pentagonal disk 30 to 35 Å thick. One side of the disk, where both the N and C termini reside, is broader than the other, giving the pentamer the appearance of a truncated pyramid. Its base has a pentagonal edge of ~42 Å (Fig. 2), which narrows to ~35 Å in the middle of the pentamer. The subunits are packed tightly around the axis of symmetry, leaving only a narrow pore through the center, with a diameter of ~5 Å in CcmL and ~3.5 Å in OrfA. This tight packing around a narrow pore is reminiscent of the hexameric carboxysome shell proteins. The tight packing of those shell subunits has been proposed to limit movement of substrates and products into and out of the carboxysome (*11*, *12*).

The database of known protein structures was searched for proteins similar to CcmL and OrfA. A single protein, EutN from *E. coli*, was identified as having a similar three-dimensional fold. The protein backbone of EutN superimposes on that of CcmL with an RMSD of only 1.3 Å (Fig. 2C). The structure of EutN had been determined as part of a structural genomics program (*21*), so detailed biological interpretation was not provided. It is known, however, that the *eut* operon, which has been studied in *Salmonella typhimurium* (*22*), encodes several proteins involved in ethanolamine utilization, which takes place inside the *eut* microcompartment. Remarkably, however, the reported EutN structure is hexameric rather than pentameric (Fig. 2). The difference between the oligomeric state of EutN, compared with those of CcmL and OrfA, presumably reflects structural differences between the *eut* microcompartment and the carboxysome, as well as a different functional role for EutN. Aside from EutN, no other homolog of CcmL or OrfA could be identified in the genome of *E. coli*, which suggests that some other potentially unrelated protein might serve as a pentameric shell protein in the *eut* microcompartment. Alternatively, the *eut* microcompartment could lack pentamers, which might explain why the microcompartments of enteric bacteria tend to have less-regular icosahedral shapes than carboxysomes.

The pentameric organization of CcmL and OrfA suggests that these proteins serve as

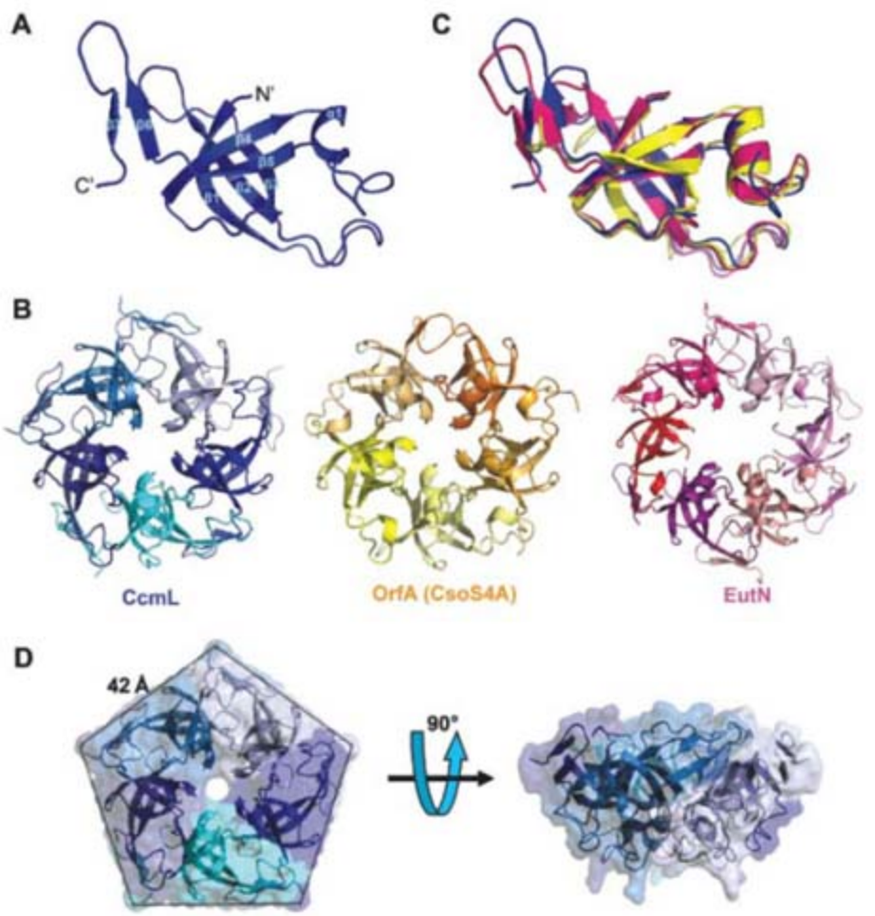


Fig. 2. Crystal structures of the carboxysome proteins CcmL and OrfA revealing pentagonal symmetry. (**A**) Structure of the CcmL monomer from *Syn.* 6803. (**B**) A comparison of similar structures: CcmL (blue), OrfA (or CsoS4A) from *H. neapolitanus* (yellow), and EutN from *E. coli* (pink) (PDB 2Z9H). The RMSD between the protein backbones of CcmL and OrfA is 1.0 Å, and 1.3 Å between CcmL and EutN. (**C**) CcmL and OrfA assemble as natural pentamers. EutN, which is part of the *eut* operon that encodes proteins presumed to comprise the distinct eut microcompartment in *E. coli*, is instead hexameric. (**D**) Top and side views of the CcmL pentamer showing a pentagonal disk with slanted sides.

vertices in their respective carboxysome shells. Whether there are significant functional differences between OrfA and its adjacent paralog in *H. neapolitanus*, OrfB, is unknown. The role proposed here for CcmL and OrfA is consistent with the earlier observation that deleting the *ccmL* gene leads to elongated, rod-shaped carboxysomes (*9*), as would be expected if the component required for vertex formation is lost. In large icosahedral shells, 12 pentamers are present among a much greater number of hexamers. In the carboxysome, only 60 copies of the CcmL or OrfA (and possibly OrfB) subunits would be expected to be present among about four or five thousand hexameric shell subunits (*3*, *12*, *23*) and a similar number of total RuBisCO subunits (*14*). This explains the inability in earlier experiments to identify OrfA or OrfB in preparations of carboxysomes.

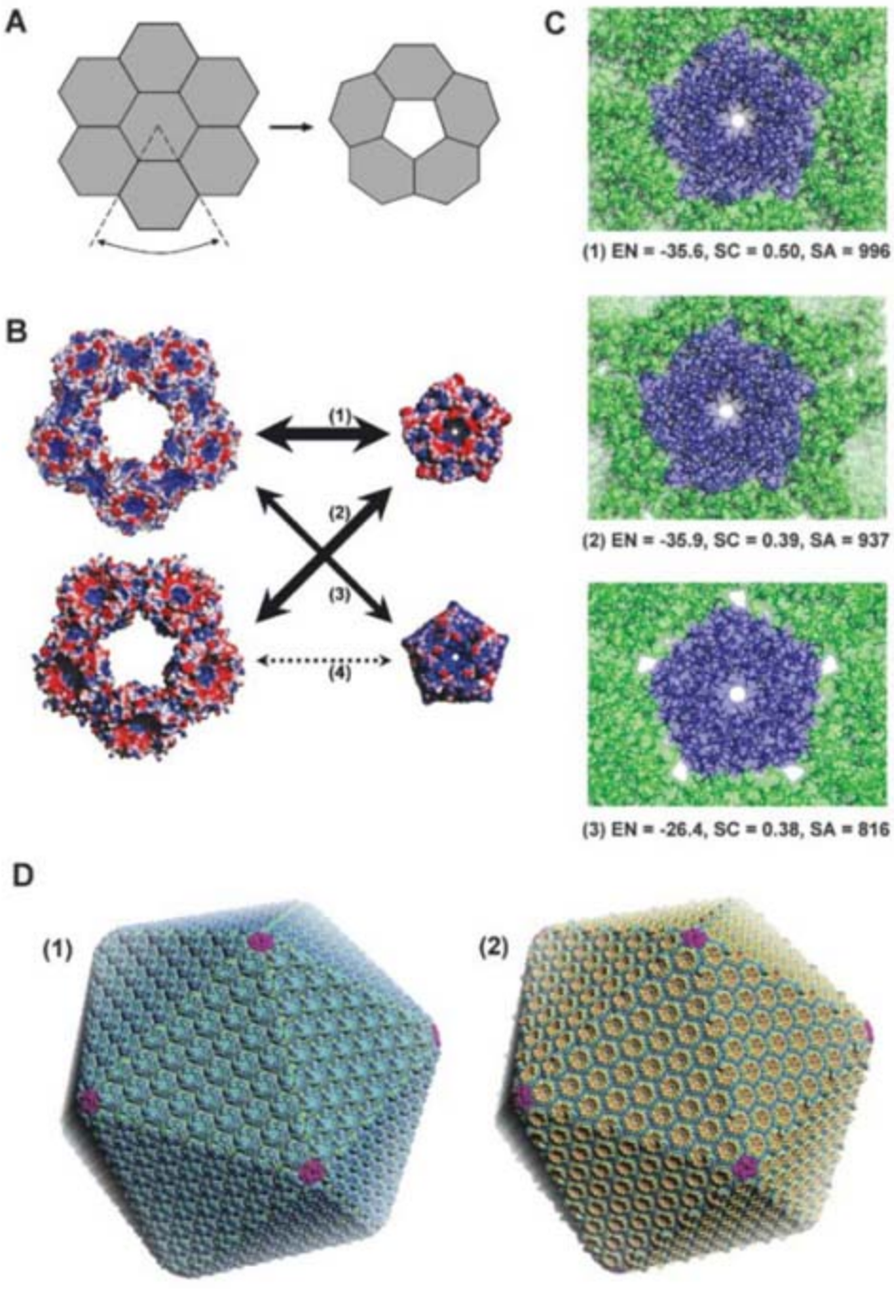


Fig. 3. Models of the carboxysome shell based on pentamer and hexamer components. (**A**) A flat layer of hexagons can be folded to give pentagonal vertices by removing one sector at each vertex. Twelve such vertices are present in an icosahedral shell. (**B**) Taken in combination, alternate choices for the curvature of the hexagonal layer and the orientation of the pentamer lead to four possible constructions, numbered 1 to 4 according to the quality of fit. Combination 4 led to impossible steric collisions. The structures are colored according to calculated electrostatic potential, from negative (red) to positive (blue). (**C**) Illustration of the best packing solutions for constructions 1 to 3. EN, calculated packing energies (*27*) (with more negative values being favorable); SC, surface complementarity (*26*); and SA, buried surface area between a pentamer and a single neighboring hexamer (with higher values of these parameters being favorable). (**D**) Two alternate models for the complete carboxysome shell, based on the two constructions, 1 and 2, judged to be most plausible. There are 740 hexamers and 12 pentamers in a T = 75 arrangement. The packing of hexamers is derived from multiple consistent crystal structures. The two models differ with respect to the orientation of the hexameric layer. The hexagonal layer is colored according to hydrophobicity, with increases showing as blue to orange. The CcmL pentamers are shown in magenta. The diameter from vertex to vertex is 1150 Å.

Knowing the structures of both the hexameric and pentameric components of the carboxysome, it is possible to model the architecture of the complete shell. The size and shape of the CcmL and OrfA pentamers make them suitable for insertion into the hexagonally packed molecular layers previously elucidated for the carboxysome shell. The hexameric units (i.e., of CcmK subunits) in the *Syn.* 6803 shell are packed together with an edge length of 40 Å (*20*) [supporting online material (SOM) text]. This size is consonant with the edge length of the CcmL pentamers. The suggestion that the CcmL and OrfA pentamers generate curvature by being inserted into an otherwise flat hexagonal molecular layer is consistent with recent electron microscopy investigations in which carboxysomes exhibit a relatively simple polyhedral shape at their vertices (*13*, *14*). Some large viral capsids have defining features, such as major protrusions, at their pentameric vertices (*24*, *25*). The lack of such protrusions at the vertices of the carboxysome places useful constraints on atomic models of the carboxysome that can be built from the known hexameric and pentameric components.

CcmL (or OrfA) pentamers can be fit into vacancies created by folding up hexameric molecular layers of CcmK (or CsoS1) proteins in four distinct ways (Fig. 3). It has not yet been possible either by biophysical methods or by electron microscopy to define which side of the hexagonally packed protein layers represents the inside of the carboxysome and which side faces outward into the bacterial cytosol (*12*). The hexagonal layers, therefore, could be folded up with either side facing outward (fig. S1). Likewise, which side of the CcmL or OrfA pentamer faces inward and which faces outward is unknown. Computational attempts were made to fit the CcmL and OrfA pentamers into their respective hexagonal layers, evaluating all four distinct (inside versus outside) orientations of the pentamers and hexamers. In addition, rotation and translation of the pentamer about its central axis was allowed, but given the tightness of the fit, only a narrow range of rotations and translations was feasible. The best-fitting solutions were identified by evaluating geometric fit (*26*) and empirical energy functions (*20*, *27*). Among the four possible orientations, one was a very poor fit and could be rejected. The other three led to potential packing solutions whose fits could be compared (fig. S2). One of these provided a much lower amount of buried surface area between the pentamer and its surrounding hexamers, and left larger gaps in the resulting shell and was, therefore, judged unlikely to be correct. Two plausible solutions for building complete models of the shell remained (Fig. 3). Both candidate

solutions orient the pentamer with its broader base facing outward. This is consistent with the role the pentamer appears to play in introducing curvature into the hexagonal layer. A notable feature of the hexagonal layer of the carboxysome shell is the presence of a bowl-shaped depression or concavity on one side of the hexameric building block (figs. S3 and S4). This depression is situated on the same side of the hexamer as both the N and C termini of the protein chain, and the prominence of the depression is affected by the disposition of the C terminus, which tends to vary in conformation between different homologs of the carboxysome hexamer (*11*, *12*). The narrow pore through the carboxysome hexamer tends to be positively charged, but the bowl-shaped depression surrounding the pore has a partial hydrophobic character, giving one side on the hexameric layer a distinctive appearance (Fig. 3). The model building does not resolve whether this side of the layer faces inward or outward. In one of the plausible models constructed here, it faces inward and could interact with RuBisCO, carbonic anhydrase, or possibly other protein components.

The present findings clarify the roles of previously uncharacterized proteins in the carboxysome; accordingly, we propose that the genes *orfA* and *orfB* be renamed *csoS4A* and *csoS4B*, consistent with other known shell protein genes in α-type carboxysomes. The results also lead to atomic models for the carboxysome shell, but these are incomplete in numerous respects, and considerably more work will be required to complete our understanding of this structure. For instance, the details of how the pentamers and hexamers fit together are only approximated here, and how the edges of the icosahedral shell are formed where a hexagonal layer bends is unknown. In addition, interactions that are almost certain to exist between the shell and the other enzymatic components of the carboxysome are only beginning to be elucidated (*28*).

The emerging structure of the carboxysome emphasizes common principles under which bacterial microcompartments and certain previously characterized viral capsids are constructed and illustrates the ability of evolution to produce a diversity of highly complex molecular structures from common principles and a small repertoire of basic building blocks. The elucidation of the component structures of the carboxysome shell also opens up strategies for designing new kinds of molecular containers or reaction chambers on the mid-nanometer scale.

References and Notes

1. J. M. Shively, R. S. English, S. H. Baker, G. C. Cannon, *Curr. Opin. Microbiol.* **4**, 301 (2001).
2. M. R. Badger, G. D. Price, *J. Exp. Bot.* **54**, 609 (2003).
3. S. Heinhorst, G. C. Cannon, J. M. Shively, in *Complex Intracellular Structures in Prokaryotes*, J. M. Shively, Ed. (Springer, Berlin, 2006), vol. 2, pp. 141–165.
4. T. A. Bobik, *Appl. Microbiol. Biotechnol.* **70**, 517 (2006).
5. J. T. Penrod, J. R. Roth, *J. Bacteriol.* **188**, 2865 (2006).
6. J. M. Shively, F. L. Ball, B. W. Kline, *J. Bacteriol.* **116**, 1405 (1973).
7. E. Gantt, S. F. Conti, *J. Bacteriol.* **97**, 1486 (1969).
8. R. S. English, S. C. Lorbach, X. Qin, J. M. Shively, *Mol. Microbiol.* **12**, 647 (1994).
9. G. D. Price, S. M. Howitt, K. Harrison, M. R. Badger, *J. Bacteriol.* **175**, 2871 (1993).
10. A. K. So *et al.*, *J. Bacteriol.* **186**, 623 (2004).
11. C. A. Kerfeld *et al.*, *Science* **309**, 936 (2005).
12. Y. Tsai *et al.*, *PLoS Biol.* **5**, e144 (2007).
13. M. F. Schmid *et al.*, *J. Mol. Biol.* **364**, 526 (2006).
14. C. V. Iancu *et al.*, *J. Mol. Biol.* **372**, 764 (2007).
15. M. R. Badger, D. T. Hanson, G. D. Price, *Funct. Plant Biol.* **29**, 161 (2002).
16. J. M. Shively *et al.*, *Can. J. Bot.* **76**, 906 (1998).
17. G. C. Cannon *et al.*, *Appl. Environ. Microbiol.* **67**, 5351 (2001).
18. J. E. Johnson, J. A. Speir, *J. Mol. Biol.* **269**, 665 (1997).
19. D. L. Caspar, A. Klug, *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1 (1962).
20. Materials and methods are available as supporting material on *Science* Online.
21. Z. Wunderlich *et al.*, *Proteins* **56**, 181 (2004).
22. E. Kofoid, C. Rappleye, I. Stojiljkovic, J. Roth, *J. Bacteriol.* **181**, 5317 (1999).
23. J. M. Shively, R. S. English, *Can. J. Bot.* **69**, 957 (1991).
24. J. T. Huiskonen, V. Manole, S. J. Butcher, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6666 (2007).
25. P. L. Stewart, R. M. Burnett, M. Cyrklaff, S. D. Fuller, *Cell* **67**, 145 (1991).
26. M. C. Lawrence, P. M. Colman, *J. Mol. Biol.* **234**, 946 (1993).
27. J. J. Gray *et al.*, *J. Mol. Biol.* **331**, 281 (2003).
28. B. M. Long, M. R. Badger, S. M. Whitney, G. D. Price, *J. Biol. Chem.* **282**, 29323 (2007).
29. See the Virus Particle Explorer, http://viperdb.scripps.edu/.

30. The authors thank D. Cascio for crystallographic advice and the staff at the Advanced Light Source beamline 8.2.2 for technical assistance. This work was supported by a grant from the Biological and Environmental Research program of the Department of Energy Office of Science. Support was also provided by NSF (grants MCB-0444568 and DMR-0213883 to G.C.C. and S.H., respectively). Protein structure coordinates and structure factors have been deposited in the Protein Data Bank (PDB); accession numbers are 2QW7 (CcmL), 2RCF (OrfA or CsoS4A), and 3BN4 (CcmK1).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1083/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Tables S1 and S2
References

9 October 2007; accepted 3 January 2008
10.1126/science.1151458

Differential Regulation of Dynein and Kinesin Motor Proteins by Tau

Ram Dixit, Jennifer L. Ross,* Yale E. Goldman, Erika L. F. Holzbaur†

Dynein and kinesin motor proteins transport cellular cargoes toward opposite ends of microtubule tracks. In neurons, microtubules are abundantly decorated with microtubule-associated proteins (MAPs) such as tau. Motor proteins thus encounter MAPs frequently along their path. To determine the effects of tau on dynein and kinesin motility, we conducted single-molecule studies of motor proteins moving along tau-decorated microtubules. Dynein tended to reverse direction, whereas kinesin tended to detach at patches of bound tau. Kinesin was inhibited at about a tenth of the tau concentration that inhibited dynein, and the microtubule-binding domain of tau was sufficient to inhibit motor activity. The differential modulation of dynein and kinesin motility suggests that MAPs can spatially regulate the balance of microtubule-dependent axonal transport.

Active transport of cytoplasmic material along microtubules is critical for cell organization and function, and defects in this process are associated with dysfunction and disease (*1*). Much of the active transport in cells depends on the molecular motor proteins cytoplasmic dynein and kinesin-1, which transport cargo toward the minus-end (toward the cell center) and plus-end of microtubules (toward the cell periphery), respectively. Dynein and kinesin have very different structures and translocation mechanisms (*2*). Kinesin has a compact motor domain and walks unidirectionally along single protofilaments with 8-nm steps (*2*). In contrast, dynein has a larger, more complex motor domain and is capable of variable step sizes, lateral steps across the microtubule surface, and processive runs toward both the minus- and plus-end of the microtubule (*3*–*5*). Cytoplasmic dynein function in vivo also requires an accessory complex, dynactin. This large, multiprotein complex is thought to facilitate dynein processivity (*6*) and may also regulate dynein activity (*5*). Within the cell, the balance between oppositely directed transport determines the steady-state distribution of organelles and biomolecules.

In the crowded cell environment, dynein and kinesin compete with nonmotile microtubule-associated proteins (MAPs) for binding to the microtubule surface. MAPs bound to microtubules might also block the path of motor proteins. Thus, MAPs can provide spatio-

Department of Physiology and Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, PA 19104, USA.

*Present address: 302 Hasbrouck Laboratory, Department of Physics, University of Massachusetts at Amherst, Amherst, MA 01003, USA.

†To whom correspondence should be addressed. E-mail: holzbaur@mail.med.upenn.edu

temporal regulation of motor proteins in vivo. Tau, a neuronal MAP, inhibits kinesin activity in vivo and in vitro (*7*–*10*); however, its effect on dynein activity is not well understood. We sought to directly observe individual encounters between single dynein or kinesin motors and tau on microtubule tracks to determine how structurally distinct motors respond to obstacles in their path.

Tau is expressed in neurons as multiple splice forms in a developmentally regulated manner (*11*). These isoforms differ in the number of microtubule-binding repeats and the length of the projection domain (Fig. 1A). We focused on the shortest and longest tau isoforms, tau23 and tau40, to compare their overall effects on motor proteins. To directly visualize tau on microtubules, we fluorescently tagged purified recombinant tau proteins with Alexa 546 dye (fig. S1A). Functionality of labeled tau was demonstrated in microtubule-binding assays (fig. S2).

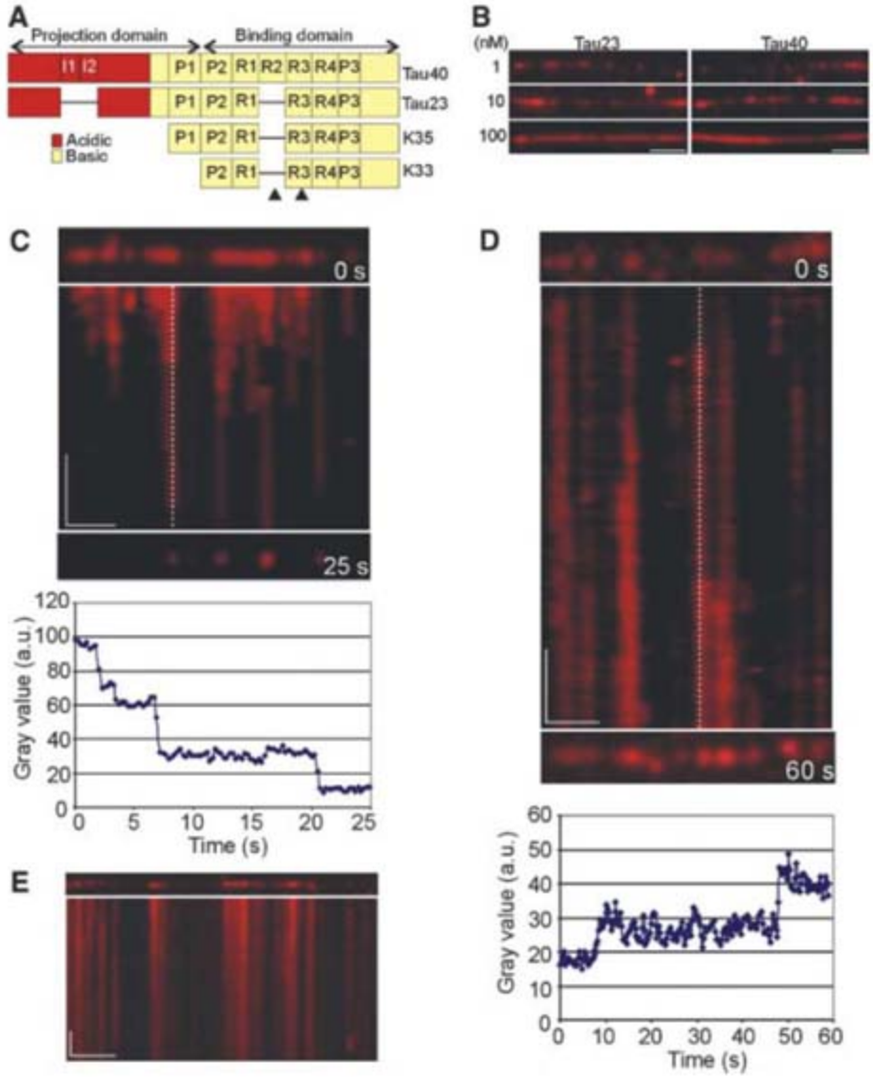


Fig. 1. Decoration of microtubules by Alexa-labeled tau. (**A**) Tau structure consists of an acidic projection domain and a basic microtubule-binding domain containing microtubule-binding repeats (R1 to R4) and proline-rich regions (P1 to P3). In the mammalian central nervous system, tau40 is the longest isoform and tau23 is the shortest isoform, which lacks insertions (I1 and I2) in the projection domain and the R2 microtubule-binding motif. K35 and K33 are recombinant variants of tau23 that are truncated at their N termini. Arrowheads indicate the position of cysteine residues used for conjugation of Alexa546. (**B**) Microtubule labeling by Alexa-tau23 and Alexa-tau40 shows a concentration-dependent increase in patch size and fluorescence intensity. Scale bars in (B) denote 2 μm. (**C**) Photobleaching of 10 nM Alexa-tau23 patches with a laser intensity 10 times as great as that used in (D) and (E) shows stepwise decreases in fluorescence intensity of the tau patches. Fluorescence intensity along the dotted line is plotted below, indicating four Alexa-tau23 molecules at this position. (**D**) Binding of 10 nM Alexa-tau23 to a microtubule shows a stepwise increase in fluorescence intensity of the tau patches. Fluorescence intensity along the dotted line is plotted below, showing sequential addition of two Alexa-tau23 molecules. (**E**) Kymograph shows stable tau decoration during the observation period. *x*-scale bar, 2 μm; *y*-scale bar, 10 s.

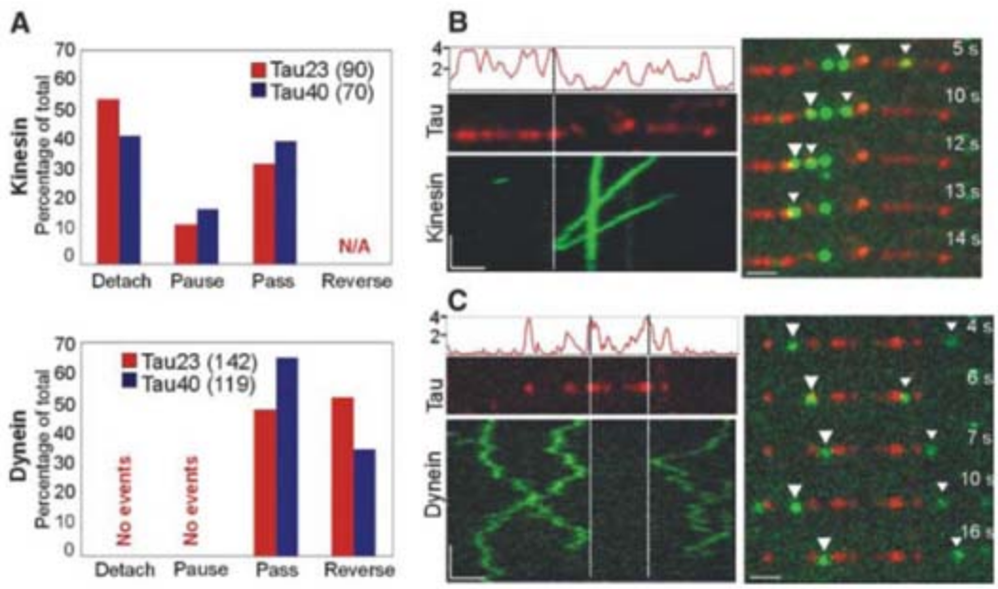


Fig. 2. Direct observation of encounters between single molecules of kinesin or dynein-dynactin and Alexa-labeled tau. (**A**) The effect of Alexa-tau23 (red bars) and Alexa-tau40 (blue bars) on kinesin and dynein motility. Kinesin either detaches, passes, or pauses (stationary for ≥ 10 frames) at a tau patch, whereas dynein either reverses direction or passes through a tau patch. The number of events is indicated in parentheses. (**B** and **C**) (Left) An Alexa-tau–decorated microtubule (red), with the relative intensity of tau fluorescence reported as the estimated number of tau molecules on the *y* axis. The kymograph (green) shows dissociation of two kinesin molecules (B) or directional reversal of two dynein-dynactin molecules (C) upon encountering a tau cluster (dotted line). (Right) Select images from each experiment. The arrowheads mark the kinesin or dynein molecules that encountered tau. *x*-scale bar, 1 μm; *y*-scale bar, 5 s. Scale bars in the right panels denote 1 μm.

Alexa-labeled tau23 and tau40 decorated Taxol-stabilized microtubules in patches whose size and fluorescence intensity increased with tau concentration (Fig. 1B). At 10 nM tau, patches were composed of 3 to 20 labeled tau molecules. Analysis of the spatial frequency spectrum of fluorescence intensity along the microtubule suggested that tau patches typically extended up to 1.2 μm (see legend to fig. S4). At peak intensity within a patch, we observed four or five bleaching events (Fig. 1C), corresponding to ~10 total tau molecules per diffraction-limited spot. These patches were formed by stepwise association of single tau molecules (Fig. 1D), consistent with tau's property of self-associating into clusters along the microtubule surface (*12*). The low–ionic strength buffer used in our experiments may fa-

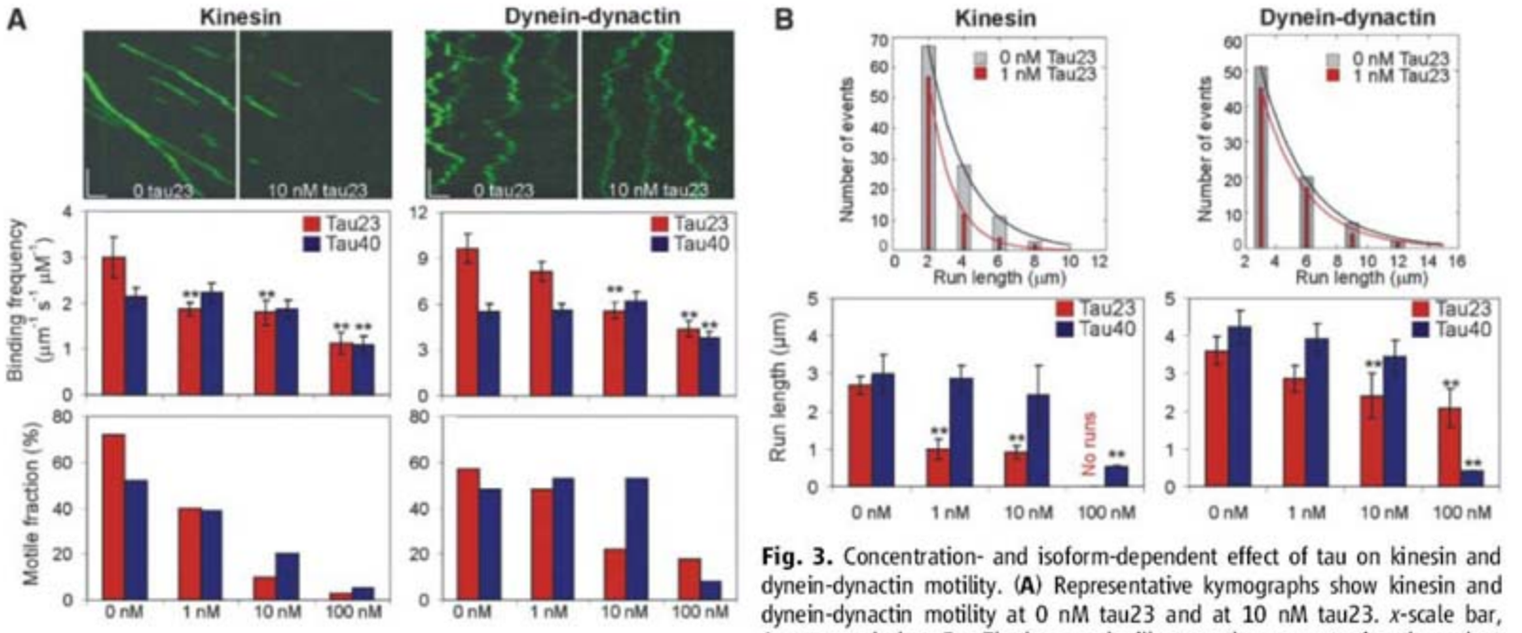
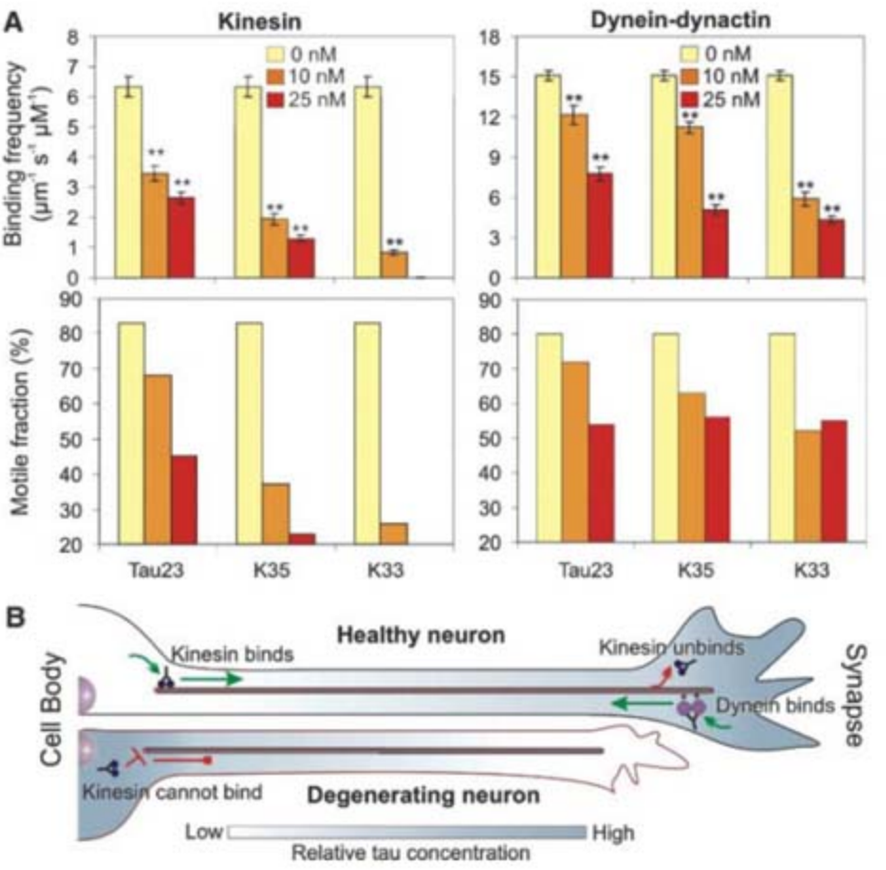


Fig. 3. Concentration- and isoform-dependent effect of tau on kinesin and dynein-dynactin motility. (**A**) Representative kymographs show kinesin and dynein-dynactin motility at 0 nM tau23 and at 10 nM tau23. *x*-scale bar, 1 μm; *y*-scale bar, 5 s. The bar graphs illustrate the concentration-dependent effect of tau23 (red bars) and tau40 (blue bars) on the average binding frequency and motile fraction of kinesin and dynein-dynactin. The error bars represent the SEM of ≥ 100 events for 0 to 10 nM tau and the SEM of ~50 events for 100 nM tau. A statistically significant difference ($P < 0.05$) from control is indicated by asterisks. (**B**) The histograms show the differential effect of 1 nM tau23 (orange bars) on the run length distribution of kinesin and dynein-dynactin. The bar graphs illustrate the concentration-dependent effects of tau23 (red bars) and tau40 (blue bars) on the average run length of kinesin and dynein-dynactin. The error bars represent the SEM of ≥ 100 events for 0 to 10 nM tau and the SEM of ~30 events for 100 nM tau. A statistically significant difference ($P < 0.05$) from control is indicated by asterisks.

Fig. 4. Comparison of the inhibitory effect of full-length and truncated versions of tau23 on kinesin and dynein-dynactin motility. (**A**) The bar graphs illustrate the effect of 0 nM (yellow bars), 10 nM (orange bars) and 25 nM (red bars) tau23 and its truncation variants, K35 and K33, on the average binding frequency and percent motile fraction of kinesin and dynein-dynactin. The error bars represent the SEM of ≥ 50 events. A statistically significant difference ($P < 0.05$) from control is indicated by asterisks. (**B**) Model of the role of tau in the regulation of axonal transport. In a healthy neuron, tau can be distributed in a proximal-distal gradient (shown in gray) that allows kinesin-driven anterograde transport from the cell body (green arrow). At the synapse, the relatively high tau concentration facilitates kinesin dissociation (red arrow). However, dynein is able to bind to distal microtubules because of its lower sensitivity to tau. In Alzheimer's disease (degenerating neuron), tau accumulates at the soma and consequently inhibits kinesin-driven anterograde transport (red blocked arrow), leading to neurodegeneration.



cilitate tau clustering, but concentration-dependent tau clustering has also been observed along axons in cultured neurons (*13*). Under the conditions of our experiments, the Alexa-tau patches were stable over the time course of several minutes (Fig. 1E), significantly longer than the average duration of motor runs (5 to 30 s). This feature allowed the direct visualization of encounters between motors and tau molecules bound to the microtubule surface.

Interactions between green fluorescent protein (GFP)-labeled dynein-dynactin isolated from mouse brain (*5*) or recombinant kinesin-GFP (*14*) (fig. S1, B and C) and Alexa-labeled tau on microtubules were observed directly by total internal reflection fluorescence microscopy (*5*). The frequency of kinesin-GFP binding to microtubules decreased with increasing local tau fluorescence intensity, whereas the frequency of dynein-dynactin-GFP binding was nearly independent of local tau fluorescence intensity (fig. S3). Thus, kinesin-GFP binding was inhibited more effectively by tau than was dynein-dynactin-GFP binding.

When single kinesin-GFP motors encountered tau patches on the microtubule, most of the motors detached from the microtubule surface (Fig. 2, A and B, and movies S1 and S2). For a given tau patch, kinesin was likely to detach at the position of peak tau fluorescence intensity (Fig. 2B). Hence, the probability of kinesin detachment is a function of the local tau concentration. Inhibition of kinesin motility was also tau-isoform specific. When kinesin-GFP did bind within a tau23 patch, mean run length was reduced by ~50% compared to mean run length along bare microtubule regions (from 1.7 ± 0.25 μm to 0.9 ± 0.03 μm; $n \geq 100$). Although tau40 patches also induced detachment of kinesin-GFP molecules (Fig. 2A and movie S2), the mean run length of kinesin-GFP moving within tau40 patches was not affected significantly (1.89 ± 0.36 μm; $n = 51$).

In contrast to kinesin, encounters of dynein-dynactin-GFP with Alexa-tau molecules did not induce motor detachment or pausing. Instead, dynein-dynactin tended to reverse direction upon encountering tau (Fig. 2, A and C, and movies S3 and S4). When dynein-dynactin bound within a tau23 patch, it did not rapidly detach like kinesin, but remained paused for relatively long periods (3 to 15 s). Tau40 also induced directional reversals as dynein-dynactin approached tau patches (Fig. 2A); however, the velocity and run length of dynein-dynactin within an Alexa-tau40 patch were not affected significantly (movie S4).

Thus, dynein-dynactin's response to encounters with tau was fundamentally different from kinesin's response, which may be related to dynein-dynactin's innate ability to execute processive backward runs. This ability would allow dynein-dynactin to remain attached to the microtubule after encounters with tau, whereas kinesin's inflexible nature makes it more likely to detach from the microtubule.

To examine the concentration-dependent effect of tau on motor function, we varied unlabeled tau23 and tau40 over a range encompassing the normal endogenous molar ratio of tubulin:tau of ~20:1. Tau decreased the binding frequency, motile fraction (defined as the proportion of landing motors that moved ≥250 nm), and run length of both kinesin-GFP and dynein-dynactin-GFP in a concentration- and isoform-dependent manner (Fig. 3, A and B; movies S5 to S10). The inhibitory effects of tau23 were less pronounced for dynein-dynactin than for kinesin. For example, 1 nM tau23 (50:1 tubulin:tau molar ratio) inhibited significantly the binding frequency and run length of kinesin but not of dynein-dynactin (Fig. 3). Significant inhibition of dynein was not observed until tau23 concentration was increased 10-fold. Tau40 was a less potent inhibitor for both motors; we did not obtain significant inhibition of either kinesin or dynein-dynactin motility at physiologically relevant tubulin:tau molar ratios with this isoform (Fig. 3). Neither tau23 nor tau40 significantly changed motor velocity of either kinesin or dynein-dynactin at any of the concentrations tested (table S1).

From a structural standpoint, tau might inhibit motor function through its projection domain and/or its microtubule-binding domain. To determine the relative contribution of these domains to motor inhibition, we studied the effects of truncated versions of tau23 on motor function. The truncated polypeptides, designated K35 and K33, differ in the length of their projection domain (Fig. 1A) but have microtubule-binding affinities similar to that of full-length tau23 (*15*). Alexa-labeled K35 and K33 decorated microtubules in a pattern similar to that observed for Alexa-labeled tau23 (fig. S4). The truncated forms of tau inhibited both kinesin-GFP and dynein-dynactin-GFP in a concentration-dependent manner (Fig. 4A). Indeed, both truncated forms were stronger inhibitors of motor proteins than was full-length tau23. Thus, the microtubule-binding domain of tau was sufficient for motor inhibition. Differences in the net charge of the projection domain between tau40 (−15), tau23 (−4), and K35 (+5) correlate well with their relative ability to inhibit binding of motor proteins, suggesting that the acidic projection domain may mimic the acidic tail of tubulin and electrostatically recruit motor proteins to the microtubule surface.

We propose a model in which tau controls the balance of microtubule-dependent axonal transport in the neuron by locally modulating motor function. In a healthy neuron, tau can be distributed in a proximal-to-distal gradient (*16*, *17*). Lower tau concentration at the cell body would allow kinesin to efficiently bind to microtubules and initiate anterograde transport of cargo, whereas higher tau concentration at the synapse would facilitate cargo release (Fig. 4B). At the same time, dynein-driven retrograde transport from the distal axon would not be impeded due to dynein's lower sensitivity to tau (Fig. 4B). Neurons can potentially fine-tune the occupancy of tau along axons by phosphorylation or dephosphorylation of sites within the microtubule-binding domain (*18*). On the basis of this model, perturbing tau distribution would predictably impair axonal transport. For example, in Alzheimer's disease, tau accumulates in the somatodendritic compartment (*19*, *20*); our model then predicts that kinesin-driven anterograde transport of new material to the axon terminal would be severely compromised, leading to neurodegeneration (*1*). The differential sensitivity of dynein and kinesin to tau thus provides a mechanism for spatiotemporal regulation of axonal transport.

References and Notes

1. E. Chevalier-Larsen, E. L. Holzbaur, *Biochim. Biophys. Acta* **1762**, 1094 (2006).
2. R. D. Vale, *Cell* **112**, 467 (2003).
3. R. Mallik, B. C. Carter, S. A. Lex, S. J. King, S. P. Gross, *Nature* **427**, 649 (2004).
4. Z. Wang, M. P. Sheetz, *Cell Struct. Funct.* **24**, 373 (1999).
5. J. L. Ross, K. Wallace, H. Shuman, Y. E. Goldman, E. L. F. Holzbaur, *Nat. Cell Biol.* **8**, 562 (2006).
6. S. J. King, T. A. Schroer, *Nat. Cell Biol.* **2**, 20 (2000).
7. A. Ebneth *et al.*, *J. Cell Biol.* **143**, 777 (1998).
8. B. Trinczek, A. Ebneth, E. M. Mandelkow, E. Mandelkow, *J. Cell Sci.* **112**, 2355 (1999).
9. A. Seitz *et al.*, *EMBO J.* **21**, 4896 (2002).
10. M. Vershinin, B. C. Carter, D. S. Razafsky, S. J. King, S. P. Gross, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 87 (2007).
11. A. Andreadis, *Prog. Mol. Subcell. Biol.* **44**, 89 (2006).
12. M. Ackmann, H. Wiech, E. Mandelkow, *J. Biol. Chem.* **275**, 30335 (2000).
13. S. Konzack, E. Thies, A. Marx, E. M. Mandelkow, E. Mandekow, *J. Neurosci.* **27**, 9916 (2007).
14. R. D. Vale *et al.*, *Nature* **380**, 451 (1996).
15. N. Gustke, B. Trinczek, J. Biernat, E. M. Mandelkow, E. Mandelkow, *Biochemistry* **33**, 9511 (1994).
16. M. Kempf, A. Clement, A. Faissner, G. Lee, R. Brandt, *J. Neurosci.* **16**, 5583 (1996).
17. M. M. Black, T. Slaughter, S. Moshiach, M. Obrocka, I. Fischer, *J. Neurosci.* **16**, 3601 (1996).
18. C. Ballatore, V. M.-Y. Lee, J. Q. Trojanowski, *Nat. Rev. Neurosci.* **8**, 663 (2007).
19. S. Khatoon, I. Grundke-Iqbal, K. Iqbal, *J. Neurochem.* **59**, 750 (1992).
20. H. Braak, E. Braak, E. M. Mandelkow, *Acta Neuropathol. (Berl.)* **87**, 554 (1994).

21. We thank E. Mandelkow for the tau constructs, R. Kudaravalli for purified kinesin, S. Shah for data analysis, and M. Ostap and members of the Holzbaur and Goldman labs for helpful comments. This work was supported by NIH grants P01-AR-051174 to the Pennsylvania Muscle Institute and GM-48661 to E.L.F.H., and NSF grant NSEC DMR04-25780 to the Penn Nano/Bio Interface Center.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1152993/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References
Movies S1 to S10

14 November 2007; accepted 3 January 2008
Published online 17 January 2008;
10.1126/science.1152993
Include this information when citing this paper.

NADP Regulates the Yeast *GAL* Induction System

P. Rajesh Kumar,[1] Yao Yu,[2] Rolf Sternglanz,[2] Stephen Albert Johnston,[3] Leemor Joshua-Tor[1]*

Transcriptional regulation of the galactose-metabolizing genes in *Saccharomyces cerevisiae* depends on three core proteins: Gal4p, the transcriptional activator that binds to upstream activating DNA sequences (UAS_{GAL}); Gal80p, a repressor that binds to the carboxyl terminus of Gal4p and inhibits transcription; and Gal3p, a cytoplasmic transducer that, upon binding galactose and adenosine 5′-triphosphate, relieves Gal80p repression. The current model of induction relies on Gal3p sequestering Gal80p in the cytoplasm. However, the rapid induction of this system implies that there is a missing factor. Our structure of Gal80p in complex with a peptide from the carboxyl-terminal activation domain of Gal4p reveals the existence of a dinucleotide that mediates the interaction between the two. Biochemical and in vivo experiments suggests that nicotinamide adenine dinucleotide phosphate (NADP) plays a key role in the initial induction event.

Saccharomyces cerevisiae senses galactose or melibiose in the surrounding medium and shuttles it into the cytoplasm. Galactose is enzymatically converted by the GAL enzymes Gal1p, Gal5p, Gal7p, and Gal10p to glucose-1-phosphate (*1*). The regulatory control of this pathway is governed by "the galactose regulon" (fig. S1). The very short induction time for *GAL* genes presents a quandary because Gal3p is localized in the cytoplasm and does not appear to enter the nucleus to physically disrupt Gal80p binding to Gal4p (*2*). Gal80p, localized to the nucleus and the cytoplasm (*2*), might therefore be sequestered in the cytoplasm upon induction, but this would require rapid shuttling of the repressor out of the nucleus, or rapid turnover of the Gal4p/Gal80p complex. It therefore appears that there is a missing link to initiate rapid induction and switch the system on. To understand the molecular mechanism of the *GAL* regulatory system, we determined the structure of *S. cerevisiae* Gal80p (ScGal80p) with the activation domain of ScGal4p.

Gal4p has a C-terminal (amino acids 768 to 881) acidic activation domain (AD), a region that is also required to bind its repressor, Gal80p (*3*–*5*). We determined the structures of $Gal80p^{S2}$:P20 and $Gal80p^{S0}$:P21 ($Gal80p^{S2}$ and $Gal80p^{S0}$ are two super-repressor mutants of ScGal80p). P21 is a 21–amino acid peptide that contains the conserved region of the C-terminal AD of Gal4p (amino acids 854 to 874). P20 is a peptide that was identified from a phage-display screen selected for Gal80p binding and was also shown to activate transcription (*6*).

The crystal structures of ScGal80p reveal a three-domain architecture with an N-terminal domain consisting of a Rossmann fold, normally associated with binding of nicotinamide adenine dinucleotide phosphate [NAD(P)] cofactors. ScGal80p does not possess a classic NAD(P) binding sequence motif, GXGXX(G/A), but a slightly altered motif, GFVGLNAA (*7*), spanning amino acids 21 to 28. The C-terminal domain consists of a large β sheet that forms an extensive dimer interface with another monomer (Fig. 1A and fig. S2). A large cleft is apparent between these two domains. A smaller, third domain, located between the N- and C-terminal domains, consists of three small β strands and a helix that resemble a set of fingers at the entrance of the

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA. [2]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY 11794–5215, USA. [3]Center for Innovations in Medicine, Biodesign Institute, School of Life Sciences, Arizona State University, Post Office Box 875901, Tempe, AZ 85287–5901, USA.

*To whom correspondence should be addressed. E-mail: leemor@cshl.edu

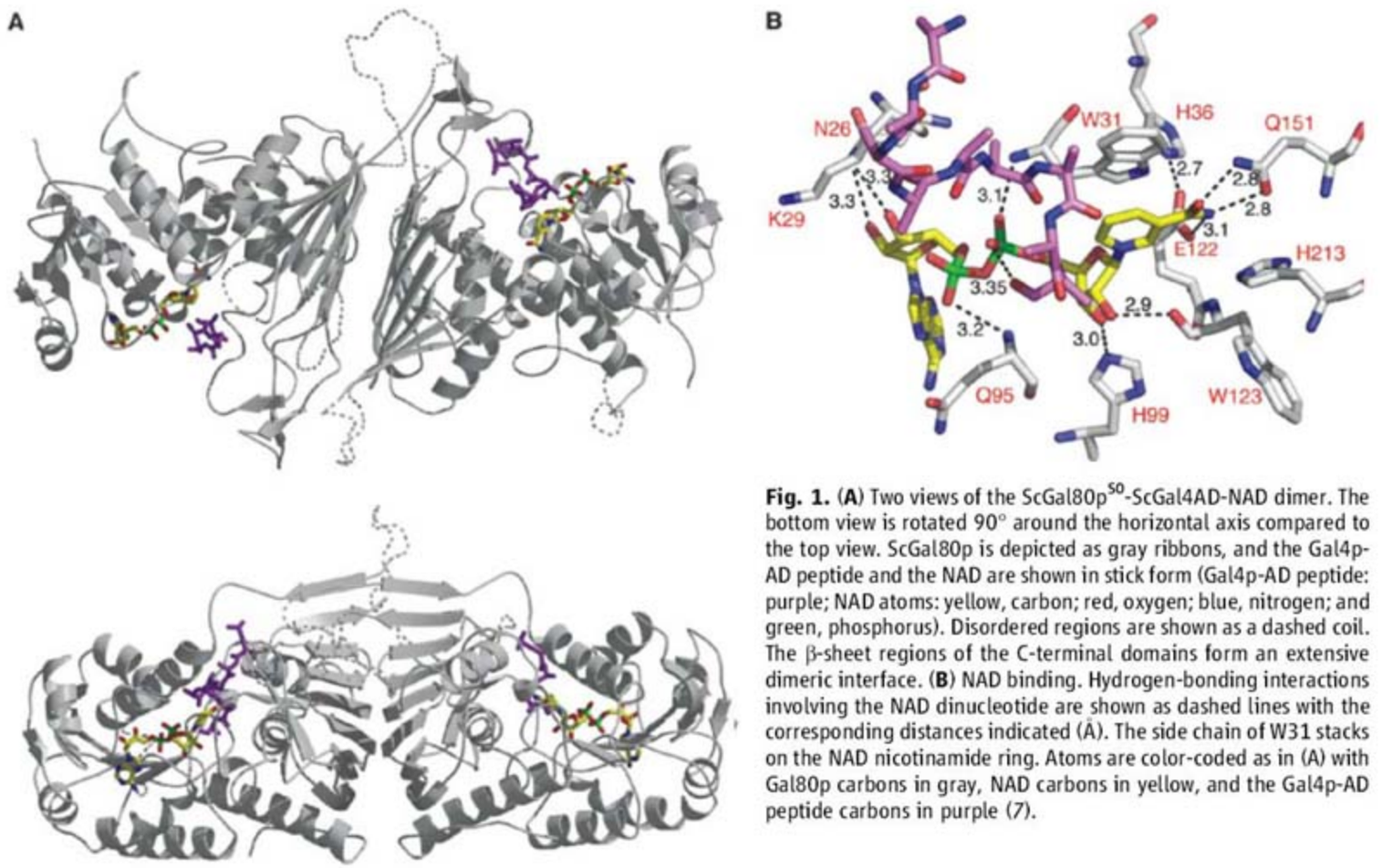


Fig. 1. (**A**) Two views of the $ScGal80p^{S0}$-ScGal4AD-NAD dimer. The bottom view is rotated 90° around the horizontal axis compared to the top view. ScGal80p is depicted as gray ribbons, and the Gal4p-AD peptide and the NAD are shown in stick form (Gal4p-AD peptide: purple; NAD atoms: yellow, carbon; red, oxygen; blue, nitrogen; and green, phosphorus). Disordered regions are shown as a dashed coil. The β-sheet regions of the C-terminal domains form an extensive dimeric interface. (**B**) NAD binding. Hydrogen-bonding interactions involving the NAD dinucleotide are shown as dashed lines with the corresponding distances indicated (Å). The side chain of W31 stacks on the NAD nicotinamide ring. Atoms are color-coded as in (A) with Gal80p carbons in gray, NAD carbons in yellow, and the Gal4p-AD peptide carbons in purple (*7*).

cleft. ScGal80p dimers form tetramers in both crystal forms (SOM text and fig S3). The structure is similar to that of Gal80p from *Kluyveromyces lactis* (KlGal80) (*8*) and closely resembles structures of several oxidoreductases.

In the ScGal80p^{S2}:P20 structure, we identified electron density that indicated an NAD dinucleotide bound to the Rossmann fold. We therefore soaked the ScGal80p^{S0}:P21 crystals, which diffracted to higher resolution, with NAD. Not only did the density of this dinucleotide become even more apparent than in the unsoaked crystals (fig. S4), but we were then able to locate a portion of the Gal4p AD peptide, which we were unable to observe previously, bound to the cleft in each monomer of Gal80p (fig. S5). We have modeled a segment of the peptide consisting of nine residues for one monomer and five residues for the other. Although the backbone electron density is clear, most side chains seem to be somewhat disordered and could not be unequivocally assigned. Nevertheless, the peptide appears to interact with the nicotinamide portion of the dinucleotide (Fig. 1). NAD nestles between Gal80p and P21, making several key interactions with ScGal80p (Fig. 1B). The crystal structure of KlGal80p did not show any bound dinucleotide (*8*).

To investigate the role of NAD as a possible cofactor of the ScGal80p-ScGal4p interaction, we performed pull-down assays (*9*) of Gal80p with purified recombinant glutathione *S*-transferase (GST)-Gal4p(768–881), containing the acidic AD, in the presence of NAD and NADH (Fig. 2A). Because Gal80p also interacts with the transducer, Gal3p, we used GST-Gal3p as a control. There was no change in binding for either of these two dinucleotides. We then tested binding in the presence of increasing concentrations of NADP and NADPH (Fig. 2A). A clear reduction in binding was observed at 5 μM NADP, and a more substantial reduction at 500 μM. NADPH also showed some reduction in Gal80p binding at 500 μM, though this could have been due to the presence of NADP impurities, which are on the order of 1 to 3%. There was no effect of Gal80p binding to GST-Gal3p with any of the four dinucleotides (Fig. 3A), indicating that this dinucleotide-regulated binding was specific to the Gal80p-Gal4p interaction. Modeling of NADP in the place of NAD in the structure did not show any clashes with either the Gal80p protein or with the Gal4p peptide.

To further analyze the effect of NADP on Gal80p binding to Gal4p-AD, we generated a panel of Gal80p mutants and tested Gal4p binding by GST pull-down experiments at varying NADP concentrations (Fig. 2B and fig. S6). Disruption of the Gal80p dimer interface (N230R) by disruption of hydrogen-bonding interactions between the monomers caused a substantial decrease in overall binding to Gal4p-AD, and this low-level binding was almost completely abolished even in the presence of low concentrations of NADP. Several mutants that should alter NAD(P) binding had lost sensitivity to NADP compared to the wild-type protein. W31A, designed to disrupt stacking with the nicotinamide ring, exhibited no sensitivity to NADP over the range of concentrations tested (0 to 2.5 mM). E122, which forms a hydrogen bond with the nicotinamide N7, was changed to an alanine and showed a decrease in overall binding but no sensitivity to NADP. H36F, which should alter positioning of E122, showed a slight decrease in binding at 2.5 mM NADP. H99A, which likely disrupts interaction of the histidine with the ribose of the nicotinamide group, behaved in a way similar to that of the wild-type protein. From modeling studies, N26 and K29 might interact with the additional 2′ phosphate of NADP. N26D showed behavior very similar to that of the wild-type protein, and K29E showed overall weaker binding.

To test the biological effect of altering the NADP binding site, we prepared Gal80p wild-type and mutants with a C-terminal FLAG tag and monitored each mutant's effect on *GAL1* expression in yeast at different time points after induction with galactose (Fig. 3 and fig. S7). All mutants predicted to affect dinucleotide binding still repressed normally in the uninduced state. Several mutants—H36F, K29E, W31A, and E122A—showed higher levels of expression at earlier time points after induction with galactose relative to the wild-type proteins. H99A was more similar to the wild type. All mutants leveled off to wild-type expression levels at 30 to 45 min after induction. This suggests that alterations in the NAD(P) binding site affects the initial rate of induction, but not overall final expression levels. It appears that NAD might facilitate Gal80p binding to Gal4p, because we could only identify Gal4p-AD with NAD bound, and NADP destabilizes this interaction. The mutations, affecting

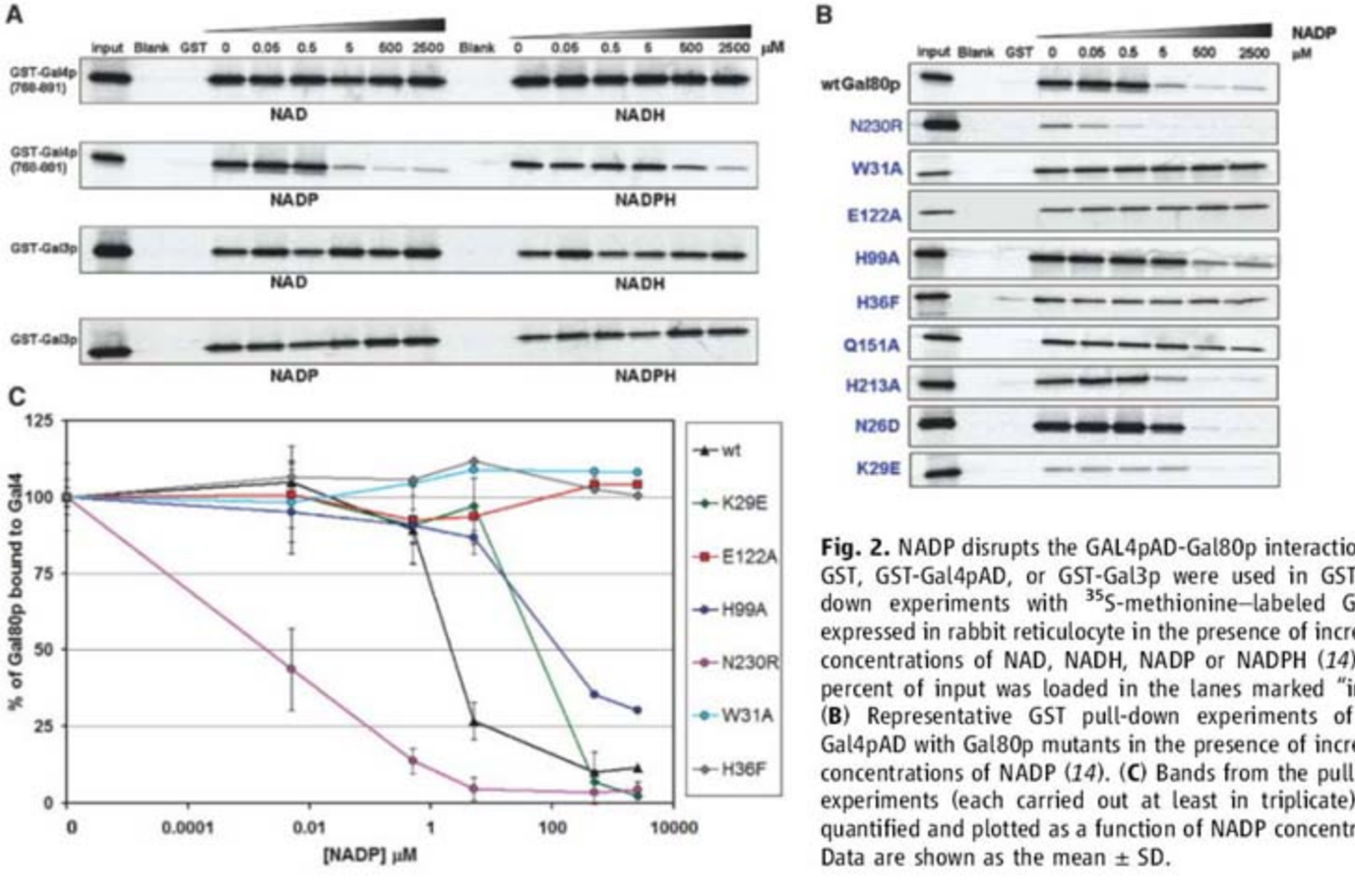


Fig. 2. NADP disrupts the GAL4pAD-Gal80p interaction. (**A**) GST, GST-Gal4pAD, or GST-Gal3p were used in GST pull-down experiments with ^{35}S-methionine–labeled Gal80p expressed in rabbit reticulocyte in the presence of increasing concentrations of NAD, NADH, NADP or NADPH (*14*). One percent of input was loaded in the lanes marked "input." (**B**) Representative GST pull-down experiments of GST-Gal4pAD with Gal80p mutants in the presence of increasing concentrations of NADP (*14*). (**C**) Bands from the pull-down experiments (each carried out at least in triplicate) were quantified and plotted as a function of NADP concentration. Data are shown as the mean ± SD.

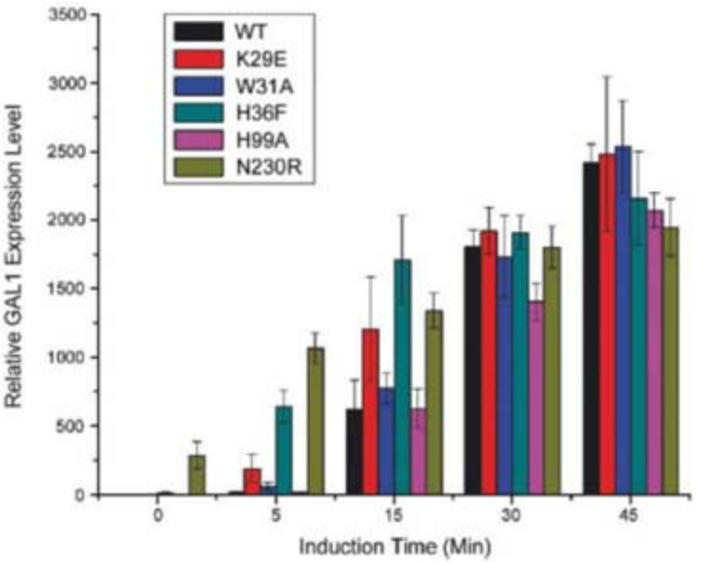


Fig. 3. Alterations in the NADP binding site changes the rate of induction in vivo. *GAL1* mRNA expression as a function of time after galactose induction. Data are shown for wild-type Gal80p and for Gal80p point mutants. All data were normalized to RNA levels measured for a control gene, *PMA1*. A *gal80*Δ mutant has a high expression level even when uninduced—as high as that seen for wild-type Gal80p when fully induced. The dimer mutant, N230R, also shows expression in the uninduced state (SOM text).

both NAD and NADP binding, would therefore disrupt both the stabilizing effect of NAD and destabilizing effect of NADP with a net result of faster induction for the mutants compared to the wild type.

The involvement of dinucleotides and metabolic factors in transcriptional regulation is seen in a few other systems. The coactivator of Oct-1, OCA-S, contains two glycolytic enzymes—glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and lactate dehydrogenase (*10*). The binding of the transcriptional corepressor complex, CtBP, is enhanced by the reduced dinucleotide NADH compared to the oxidized form (*11*) and it possesses a NAD-dependent dehydrogenase activity (*12*). The DNA-binding activity of the transcription factor neuronal PAS domain protein 2 (NPAS2) is sensitive to the oxidation state of NAD, with DNA binding enhanced by the reduced form of the dinucleotide (*13*). Although we do not understand precisely how this trigger for *GAL* regulation functions, nor the involvement of NADP versus NAD, we speculate that switching the cell to a fermentable galactose medium causes a change in NADP/NADPH or NADP/NAD ratios in the cell, and Gal80p effectively senses the metabolic state of the cell. NADP might be acting as a "second messenger" in triggering the system. Alternatively, Gal80p may function as an oxidoreductase, actively converting NADPH to NADP in the presence of a substrate and causing it to disassociate from Gal4p.

References and Notes

1. P. J. Bhat, T. V. S. Murthy, *Mol. Microbiol.* **40**, 1059 (2001).
2. G. Peng, J. E. Hopper, *Mol. Cell. Biol.* **20**, 5140 (2000).
3. J. Ma, M. Ptashne, *Cell* **50**, 137 (1987).
4. L. Keegan, G. Gill, M. Ptashne, *Science* **231**, 699 (1986).
5. S. A. Johnston, J. M. Salmeron Jr., S. S. Dincher, *Cell* **50**, 143 (1987).
6. Y. Han, T. Kodadek, *J. Biol. Chem.* **275**, 14979 (2000).
7. Abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; K, Lys; L, Leu; N, Asn; Q, Gln; R, Arg; V, Val; and W, Trp.
8. J. B. Thoden, C. A. Sellick, R. J. Reece, H. M. Holden, *J. Biol. Chem.* **282**, 1534 (2007).
9. M. B. Einarson, in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2001), pp. 18.55–18.59.
10. L. Zheng, R. G. Roeder, Y. Luo, *Cell* **114**, 255 (2003).
11. C. C. Fjeld, W. T. Birdsong, R. H. Goodman, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9202 (2003).
12. V. Kumar *et al.*, *Mol. Cell* **10**, 857 (2002).
13. J. Rutter, M. Reick, L. C. Wu, S. L. McKnight, *Science* **293**, 510 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. We thank T. Messick, K. Siddiqui, M. Rossman, J. Hicks, A. Gann, S. Harrison, and members of the Joshua-Tor lab for discussions and advice; A. Heroux, M. Becker, and H. Robinson for support at the National Synchrotron Light Source (NSLS) and S. Ginell and N. Duke for support at the Advanced Photon Source (APS). The NSLS and the APS are supported by the U.S. Department of Energy, Office of Basic Energy Sciences. Coordinates and structure factors have been submitted to the Protein Data Bank. Accession numbers: Gal80p^{S0}-ScGal4AD-NAD: 3BTS; Gal80p^{S2}: 3BTU; Gal80p^{S0}: 3BTV. This work was supported by NIH grants GM074075 (to L.J.) and GM55641 (to R.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1090/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S5
References

18 October 2007; accepted 4 January 2008
10.1126/science.1151903

A Shared Docking Motif in TRF1 and TRF2 Used for Differential Recruitment of Telomeric Proteins

Yong Chen,[1] Yuting Yang,[1] Megan van Overbeek,[2]* Jill R. Donigian,[2]* Paul Baciu,[1] Titia de Lange,[2] Ming Lei[1]†

Mammalian telomeres are protected by a six-protein complex: shelterin. Shelterin contains two closely related proteins (TRF1 and TRF2), which recruit various proteins to telomeres. We dissect the interactions of TRF1 and TRF2 with their shared binding partner (TIN2) and other shelterin accessory factors. TRF1 recognizes TIN2 using a conserved molecular surface in its TRF homology (TRFH) domain. However, this same surface does not act as a TIN2 binding site in TRF2, and TIN2 binding to TRF2 is mediated by a region outside the TRFH domain. Instead, the TRFH docking site of TRF2 binds a shelterin accessory factor (Apollo), which does not interact with the TRFH domain of TRF1. Conversely, the TRFH domain of TRF1, but not of TRF2, interacts with another shelterin-associated factor: PinX1.

Shelterin acts in conjunction with many associated factors (*1*–*6*). Most of the shelterin-associated proteins are recruited to telomeres through interactions with TRF1 or TRF2 (*2*–*6*). However, the molecular mechanism of these TRF1- and TRF2-mediated interactions remains unknown. TRF1 and TRF2 share the same molecular architecture, characterized by a C-terminal Myb/SANT DNA binding domain (*7*, *8*) and an N-terminal TRFH domain (*9*). The TRFH domains ($TRF1_{TRFH}$ and $TRF2_{TRFH}$) mediate homodimerization and are required for telomeric DNA binding by TRF1 and TRF2 (*10*, *11*). Several different protein interactions have been mapped to the TRFH domains of TRF1 and TRF2 (*2*, *12*–*14*). The TRFH domains have almost identical three-dimensional structures (*11*); therefore, it is difficult to explain how TRF1 and TRF2 can interact with different proteins.

TRF1 and TRF2 both bind to another shelterin protein: TIN2 (*12*, *15*, *16*). The TRF1-TIN2 interaction was mediated by $TRF1_{TRFH}$ and the C terminus of TIN2 (*12*). Further mapping revealed that a peptide of TIN2—denoted

[1]Department of Biological Chemistry, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, MI 48109, USA. [2]Laboratory for Cell Biology and Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: leim@umich.edu

as $TIN2_{256–276}$ [$TIN2_{TBM}$: TIN2–TRFH binding motif (TBM)]—retains the $TRF1_{TRFH}$ binding activity with a binding affinity of 314 nM (figs. S1 and S2 and Fig. 1A). To understand how $TIN2_{TBM}$ is recognized by $TRF1_{TRFH}$, we crystallized the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex and solved its structure at 2.0 Å resolution (table S1) (*17*). The electron density map shows that residues 257 to 268 of $TIN2_{TBM}$ assume a well-defined conformation (fig. S3). $TRF1_{TRFH}$ forms homodimers, and each $TRF1_{TRFH}$ interacts with one $TIN2_{TBM}$ peptide (Fig. 1B). $TRF1_{TRFH}$ exhibits essentially the same conformation as unliganded $TRF1_{TRFH}$ except for loop L34 (Fig. 1C) (*11*). Loop L34 is partially disordered

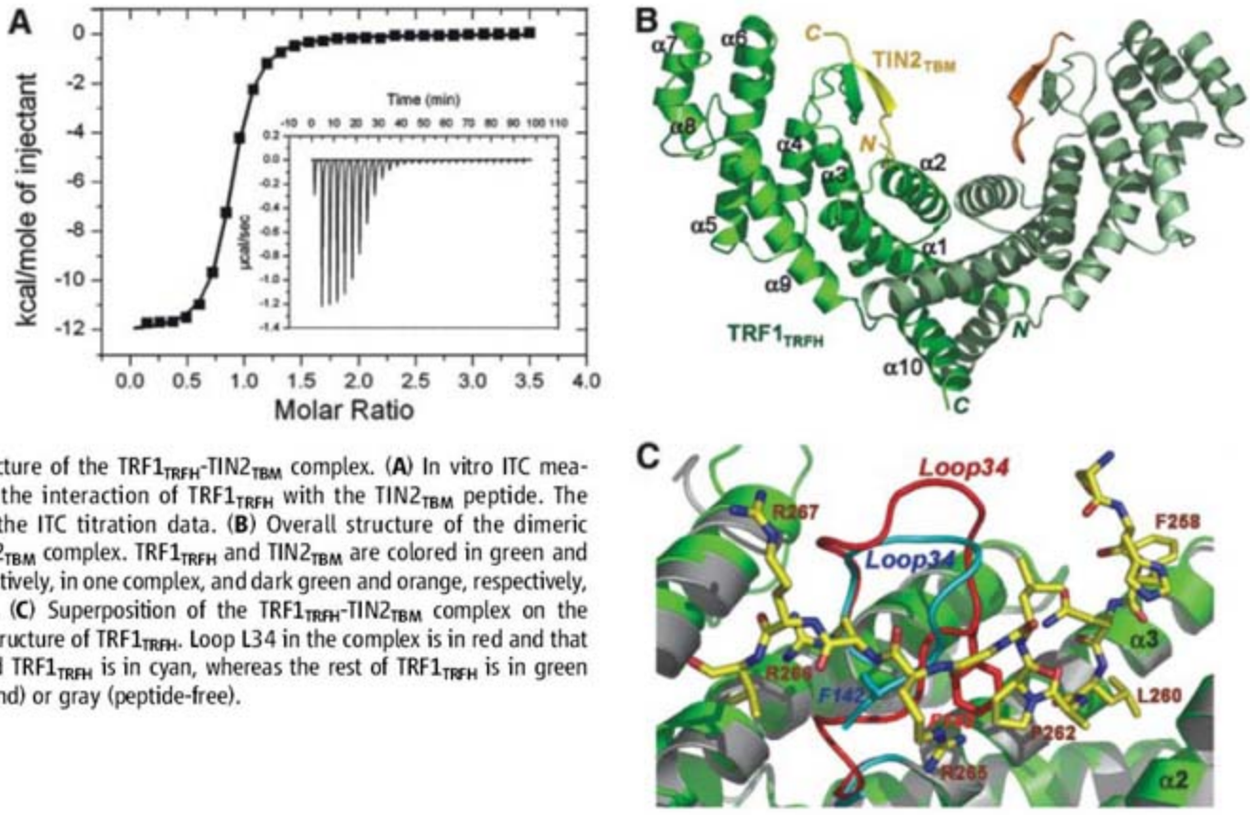
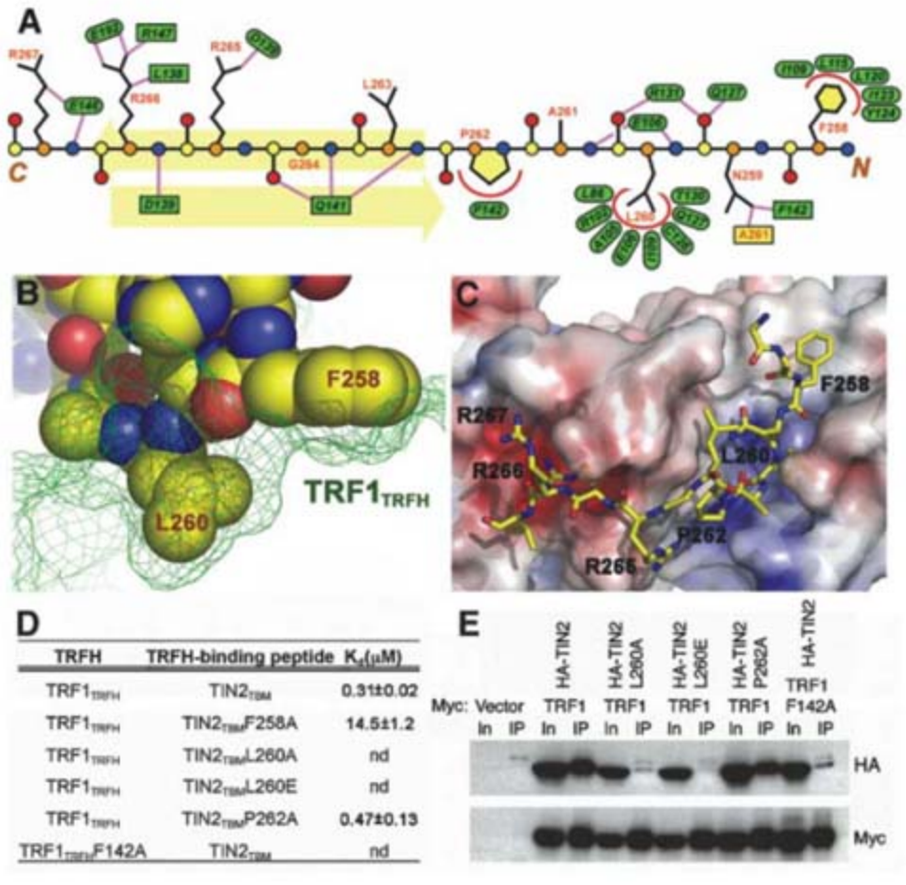


Fig. 1. Structure of the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex. (**A**) In vitro ITC measurement of the interaction of $TRF1_{TRFH}$ with the $TIN2_{TBM}$ peptide. The inset shows the ITC titration data. (**B**) Overall structure of the dimeric $TRF1_{TRFH}$-$TIN2_{TBM}$ complex. $TRF1_{TRFH}$ and $TIN2_{TBM}$ are colored in green and yellow, respectively, in one complex, and dark green and orange, respectively, in the other. (**C**) Superposition of the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex on the unliganded structure of $TRF1_{TRFH}$. Loop L34 in the complex is in red and that of unliganded $TRF1_{TRFH}$ is in cyan, whereas the rest of $TRF1_{TRFH}$ is in green ($TIN2_{TBM}$-bound) or gray (peptide-free).



TRFH	TRFH-binding peptide	K_d(µM)
$TRF1_{TRFH}$	$TIN2_{TBM}$	0.31±0.02
$TRF1_{TRFH}$	$TIN2_{TBM}$F258A	14.5±1.2
$TRF1_{TRFH}$	$TIN2_{TBM}$L260A	nd
$TRF1_{TRFH}$	$TIN2_{TBM}$L260E	nd
$TRF1_{TRFH}$	$TIN2_{TBM}$P262A	0.47±0.13
$TRF1_{TRFH}$F142A	$TIN2_{TBM}$	nd

Fig. 2. The $TRF1_{TRFH}$-$TIN2_{TBM}$ interface. (**A**) Schematic depiction of the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction. The main-chain atoms of $TIN2_{TBM}$ are shown as circles [carbon in yellow (Cα in orange), oxygen in red, and nitrogen in blue]. Residues of $TRF1_{TRFH}$ are shown as green ovals (side-chain interaction) and square boxes (main-chain interaction). Hydrophilic and hydrophobic interactions are shown as straight magenta lines and curved red lines, respectively. The pale yellow arrows denote the intermolecular β sheet. (**B**) The shape of the hydrophobic pocket of TRF1 (green mesh) complements the side chain of TIN2-L260 well. (**C**) Electrostatic surface potential of the $TIN2_{TBM}$ binding site of $TRF1_{TRFH}$. Positive potential, blue; negative potential, red. (**D**) In vitro ITC binding data of wild-type and mutant $TRF1_{TRFH}$-$TIN2_{TBM}$ interactions. K_d, equilibrium dissociation constant; nd, not detectable by ITC. (**E**) Co-IP of the same sets of mutant TRF1-TIN2 interactions (except the TRF1-TIN2 F258A interaction) as in (D). Lanes marked "In" represent 2.5% of input cell lysate used for the immunoprecipitation.

in the peptide-free structure (Fig. 1C). However, once $TIN2_{TBM}$ is bound, loop L34 folds back upon helices α3 and α4, sandwiched between the helices and $TIN2_{TBM}$ (Fig. 1C).

The structure of the complex reveals two adjacent but structurally distinct interaction modes. The N terminus of $TIN2_{TBM}$ [His^{257}-Phe-Asn-Leu-Ala-Pro^{262} (H257-F-N-L-A-P262)] (*18*) adopts an extended conformation stabilized by an extensive intermolecular hydrogen-bonding network (Fig. 2A and fig. S4). The side chain of L260 is therefore positioned into a deep hydrophobic pocket of $TRF1_{TRFH}$ (Fig. 2, B and C). In addition, F258 and P262 also make hydrophobic contacts with $TRF1_{TRFH}$: F258 sits on a concave hydrophobic surface, whereas P262 stacks with TRF1-*F142* (italics are used here for individual residues of TRF1 and TRF2) (Fig. 2, A to C, and fig. S4). In contrast, the C terminus of $TIN2_{TBM}$ (L263-G-R-R-R-V268) is positioned on the surface of loop L34 through formation of an antiparallel β sheet with *D139-A-Q141* of $TRF1_{TRFH}$ (Fig. 2A and fig. S4) so that R265-R-R267 of $TIN2_{TBM}$ contacts $TRF1_{TRFH}$ through electrostatic interactions (Fig. 2C). In particular, R266 is nested within an acidic depression on the surface of loop L34 through a network of salt bridges and hydrogen bonds (Fig. 2, A and C, and fig. S4).

To investigate the importance of the TRF1-TIN2 interaction, we first measured the binding of different mutant $TIN2_{TBM}$ peptides to $TRF1_{TRFH}$ by isothermal titration calorimetry (ITC). Substitution of L260 with either an alanine or a glutamate abolished the binding (Fig. 2D). Similarly, mutant $TIN2_{TBM}$–F258→A258 ($TIN2_{TBM}$-F258A) substantially impaired the interaction (Fig. 2D). By contrast, mutant $TIN2_{TBM}$-P262A, designed to eliminate a stacking interaction with TRF1-*F142*, had a wild-type binding affinity, indicating that loss of this interaction is not essential for binding (Fig. 2D). However, substitution of TRF1-*F142* with an alanine completely abrogated the binding to $TIN2_{TBM}$ (Fig. 2D). We then tested the interactions of mutant proteins transiently expressed in human embryonic kidney 293T cells, and the coimmunoprecipitation (Co-IP) results are consistent with the in vitro ITC measurements (Fig. 2E). We therefore conclude that the TRFH interaction motif in TRF1 is necessary for the TRF1-TIN2 interaction both in vitro and in vivo.

Given the sequence and structural similarities of the TRFH domains of TRF1 and TRF2, we expected that TRF2 would also bind to TIN2 through the TRFH domain (figs. S5 and S6). However, Co-IP studies of a specific mutant in TRF2 (TRF2-*F120A*, where TRF2-*F120* is structurally equivalent to TRF1-*F142*), which was predicted to abolish TIN2 binding to $TRF2_{TRFH}$, did not have the expected effect (Fig. 3A). Therefore, $TRF2_{TRFH}$ is not required for the stable association with TIN2 in vivo. In order to define the actual TIN2 binding site, we tested an array of glutathione *S*-transferase–TRF2 fusion fragments in a Far-Western assay for their ability to interact with TIN2. The result showed that a short peptide of TRF2 ($TRF2_{352–365}$) can mediate an efficient interaction with TIN2 (Fig. 3B). In addition, purified $TRF2_{350–366}$ comigrated with $TIN2_{1–220}$ in gel-filtration chromatographic analysis, indicating that $TIN2_{1–220}$ is sufficient for binding (fig. S7). Furthermore, Co-IP data showed that a deletion mutant of TRF2 (TRF2-Δ352–367) that retains the entire TRFH domain but lacks the TIN2 binding site failed to associate with TIN2 (Fig. 3A). Therefore, $TRF2_{TRFH}$ does not mediate a stable interaction with TIN2 in vivo. Collectively, we conclude that, although TRF1 binds TIN2 through its TRFH domain, TRF2 interacts with TIN2 through a short motif in its C terminus.

The distinctive specificity of the TRFH domains of TRF1 and TRF2 suggested that subtle structural differences are responsible for the ability of TIN2 to distinguish between these two paralogous proteins. ITC measurement showed that $TRF2_{TRFH}$ interacts with $TIN2_{TBM}$ in vitro, but with a much lower affinity (6.49 μM) (fig. S8A). To understand this binding specificity, we solved the crystal structure of the $TRF2_{TRFH}$-$TIN2_{TBM}$ complex at 2.15 Å resolution (fig. S8B and table S1). Although the overall conformations of $TIN2_{TBM}$

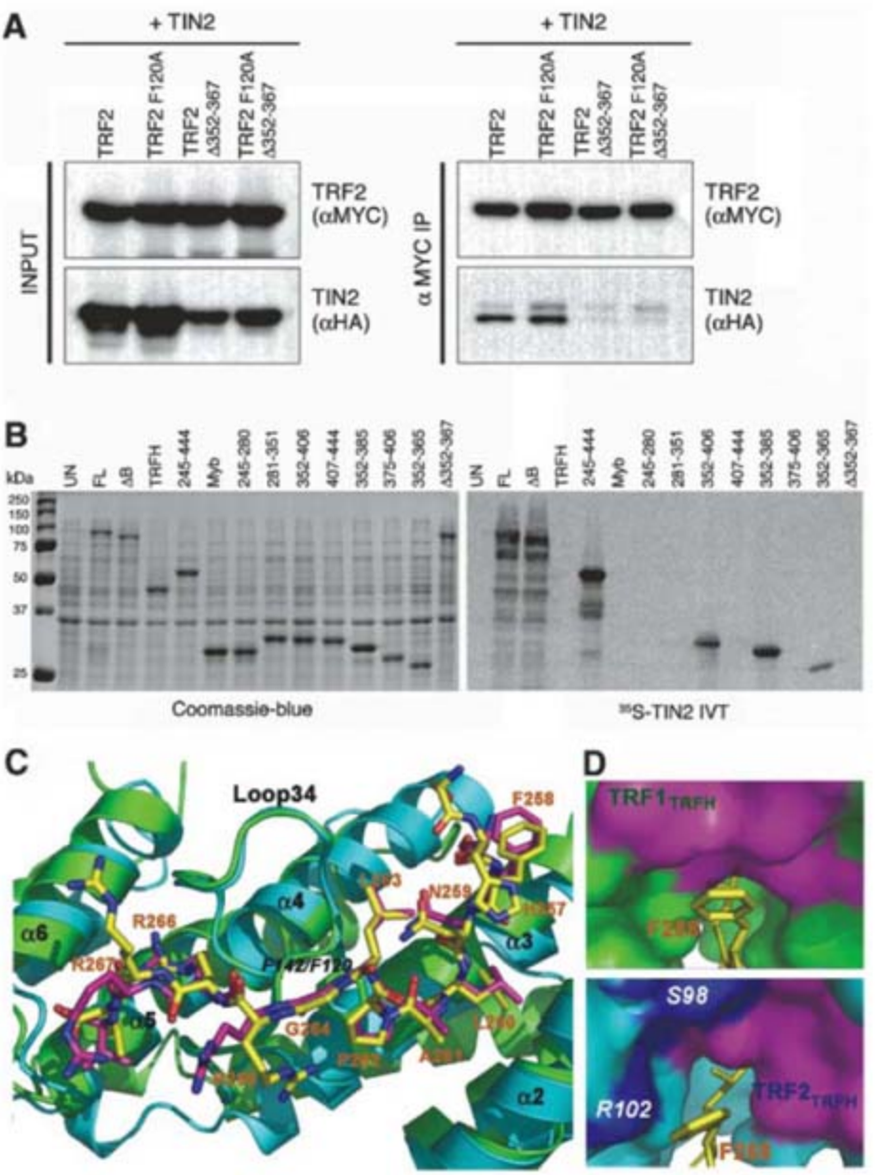


Fig. 3. The TRF2-TIN2 interaction. (**A**) Co-IP of TIN2 with cotransfected wild-type and mutant TRF2. (**B**) Far-Western analysis of the TIN2 binding region of TRF2 (FL, full-length; TRF2-ΔB, TRF2-Δ1–42). (**C**) Superposition of the $TIN2_{TBM}$ binding sites in the $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$TIN2_{TBM}$ complexes. $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are in green and cyan, respectively. The $TIN2_{TBM}$ peptides bound to $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are shown in stick model format and in yellow and magenta, respectively. (**D**) TIN2-F258 interacts less efficiently with TRF2 than with TRF1. The F258 binding surfaces of $TRF1_{TRFH}$ (top panel) and $TRF2_{TRFH}$ (bottom panel) are shown in magenta (hydrophobic patch) and blue (hydrophilic patch). The rest of $TRF1_{TRFH}$ and $TRF2_{TRFH}$ is in green and cyan, respectively.

bound to $TRF1_{TRFH}$ and $TRF2_{TRFH}$ are very similar (Fig. 3C), subtle differences can explain the difference in affinities of the two complexes (Fig. 3D and fig. S8, C and D). In the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex, TIN2-F258 sits snugly on a hydrophobic surface of $TRF1_{TRFH}$ (Fig. 3D). In contrast, F258 rotates away from the interface and packs less efficiently with $TRF2_{TRFH}$, because the edge of the interaction surface is partially occupied by polar residues *S98* and *R102* (Fig. 3D). In addition, TRF1-*E192*, which is key for $TIN2_{TBM}$ binding, is replaced by a lysine residue in TRF2 (*K173*), resulting in loss of two ion-pairing interactions and an electrostatically unfavorable contact between TIN2-R266 and TRF2-*K173* (figs. S6 and S9).

These results suggested that TRF2 might use its TRFH domain peptide docking site to recruit one or more of the shelterin accessory factors (*2*, *13*, *19*). $TRF2_{TRFH}$ is known to interact with Apollo, which functions together with TRF2 in protecting telomeres during S phase (*2*, *13*). $TRF2_{TRFH}$ directly binds to the C terminus of Apollo ($Apollo_{496–532}$) (fig. S10A) (*13*). We confirmed this interaction using the ITC binding assay (Fig. 4A). Under the same conditions, no binding enthalpy was measurable between $Apollo_{496–532}$ and $TRF1_{TRFH}$, indicating that $Apollo_{496–532}$ binding is specific for TRF2 (Fig. 4A). To understand how TRF2 recognizes Apollo, we determined the crystal structure of the $TRF2_{TRFH}$-$Apollo_{496–532}$ complex at 2.5 Å resolution (Fig. 4B and table S1). The structure clearly shows electron density corresponding to the 12 N-terminal residues of $Apollo_{496–532}$ (amino acids 498 to 509), referred to as $Apollo_{TBM}$ (fig. S10, A and B). The structure reveals that $Apollo_{TBM}$ interacts with $TRF2_{TRFH}$ through the same molecular surface as in the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex (Fig. 4C). Overlay of the two complexes reveals many similarities between the C terminus of $Apollo_{TBM}$ (Y504-L-L-T-P-V509) and the N terminus of $TIN2_{TBM}$ (F258-N-L-A-P-G265). First, two peptides are almost identical in overall conformation (Fig. 4C and fig. S10, C and D). Second, most of the hydrogen bonds in the $TRF2_{TRFH}$-$Apollo_{TBM}$ complex are conserved in $TRF1_{TRFH}$-$TIN2_{TBM}$ (fig. S10, C and E). Third, L506 and P508 of Apollo interact with $TRF2_{TRFH}$ in the same fashion as do their counterparts of $TIN2_{TBM}$ (Fig. 4C and fig. S10D). It is noteworthy that the TBMs of TIN2 and Apollo share the sequence Y/F-X-L-X-P (where X is any amino acid).

Despite the high degree of similarity between the $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$Apollo_{TBM}$ interactions, substantial structural variations are evident outside the Y/F-X-L-X-P motif. Unlike $TIN2_{TBM}$, the Y-X-L-X-P motif resides at the C terminus of $Apollo_{TBM}$, and $Apollo_{TBM}$ lacks a C-terminal polyarginine tail (Fig. 4C). Instead, it has a six-residue extension preceding the Y/F-X-L-X-P motif, which adopts a short helical conformation (Fig. 4, C and D) and packs on loop L23 and helices α2 and α3 of $TRF2_{TRFH}$ through hydrophobic contacts (Fig. 4D and fig. S10, C and E). Apollo-Y504 rotates ~ 90° relative to TIN2-F258 in the $TRF1_{TRFH}$-$TIN2_{TBM}$ complex to fit into a hydrophobic cleft formed by *L101* and *R102* of TRF2 (Fig. 4D). This reorientation of Y504 is coupled with a partial refolding of loop L23 of TRF2: TRF2-*E94* rotates ~ 180° relative to its position in the peptide-free conformation and makes two electrostatic interactions with K503 and Y504 of Apollo (Fig. 4D and fig. S10E). These marked conformational differences suggest that a tyrosine residue is preferred at the N-terminal position of the F/Y-X-L-X-P motif for efficient binding to $TRF2_{TRFH}$, whereas a phenylalanine is preferred for $TRF1_{TRFH}$. Furthermore, superposition of the $TRF1_{TRFH}$-$TIN2_{TBM}$ and the $TRF2_{TRFH}$-$Apollo_{TBM}$ complexes shows that the space occupied by L500 and Y504 of $Apollo_{TBM}$ is occluded in $TRF1_{TRFH}$, which explains why $Apollo_{TBM}$ binding is $TRF2_{TRFH}$-specific (Fig. 4E and fig. S11). Given the close similarity of the TRFH do-

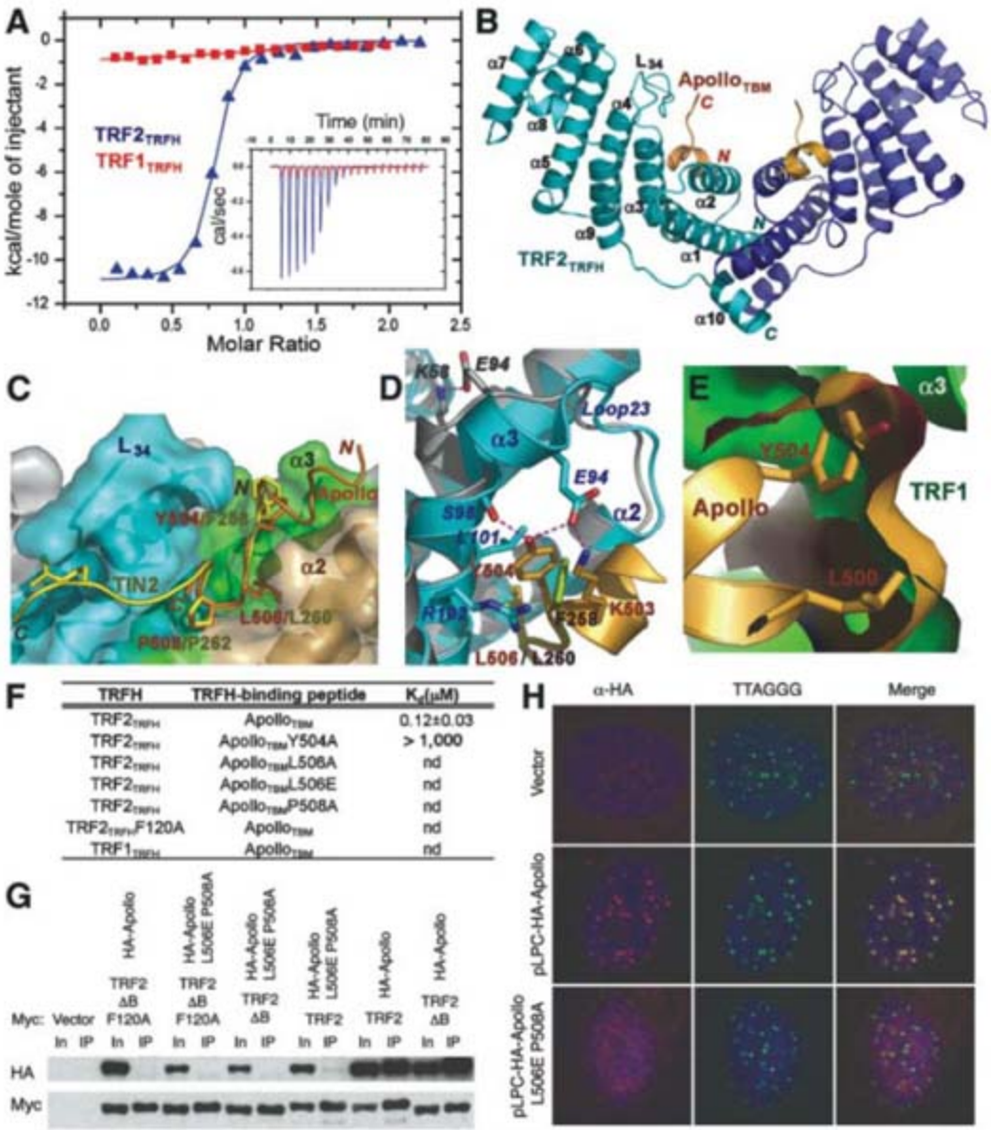


Fig. 4. The TRF2-Apollo interaction. (**A**) ITC measurement of the interactions of $TRF1_{TRFH}$ (red) and $TRF2_{TRFH}$ (blue) with the $Apollo_{TBM}$ peptide. (**B**) Overall structure of the dimeric $TRF2_{TRFH}$-$Apollo_{TBM}$ complex. (**C**) Superposition of $Apollo_{TBM}$ (orange) and $TIN2_{TBM}$ (yellow) reveals a shared F/Y-X-L-X-P motif. (**D**) Superposition of the $TRF2_{TRFH}$-$Apollo_{TBM}$ and the $TRF2_{TRFH}$-$TIN2_{TBM}$ complexes in the vicinity of the Apollo helix. The $TRF2_{TRFH}$ molecules are colored in cyan ($Apollo_{TBM}$-bound) and gray ($TIN2_{TBM}$-bound), respectively. (**E**) $Apollo_{TBM}$ binding is $TRF2_{TRFH}$-specific. The surface representations show that there is no room for Apollo L500 and Y504 to fit into the peptide binding site of $TRF1_{TRFH}$. (**F**) In vitro ITC binding data of wild-type and mutant $TRF2_{TRFH}$-$Apollo_{TBM}$ interactions. (**G**) Co-IP data show that Apollo double-mutant L504E/P506 and TRF2 single-mutant *F120A* disrupt the in vivo TRF2-Apollo interaction. (**H**) Localization of retrovirally expressed HA-tagged wild type and L506E/P508A double mutant of Apollo in BJ-hTERT cells.

mains of TRF1 and TRF2, these structural variations emphasize that the TRFH domain is a versatile framework for interactions with different proteins.

The crystal structure of the $TRF2_{TRFH}$-$Apollo_{TBM}$ complex is corroborated by mutagenesis. Mutations of the conserved hydrophobic residues of Apollo (F504, L506, and P508) or TRF2 (*F120*) completely abolished the interaction both in vitro and in vivo (Fig. 4, F and G). We further assayed the cellular localization of wild-type and mutant Apollo by expressing hemagglutinin (HA)–tagged proteins in human telomerase reverse transcriptase (hTERT)–immortalized human BJ fibroblasts. Although wild-type Apollo showed the expected telomere localization, the L506E/P508A double mutant was distributed throughout the nucleoplasm with no obvious accumulation at telomeres (Fig. 4H). This result confirms the structural information and indicates that the binding of Apollo to the TRFH domain of TRF2 is required for the telomeric localization of Apollo.

We next asked whether other shelterin-associated proteins might contain the F/Y-X-L-X-P motif suggestive of an interaction with the TRFH domain of TRF1 or TRF2. We identified this motif in PinX1, originally identified as a TRF1-interacting protein in a yeast two-hybrid screen (*6*). An 11-residue fragment of PinX1 (R287-D-F-T-L-K-P-K-K-R-R297), referred to as $PinX1_{TBM}$, closely resembles $TIN2_{TBM}$ (fig. S12A), suggesting that it may bind to $TRF1_{TRFH}$ in the same fashion as does $TIN2_{TBM}$. ITC data confirmed the $TRF1_{TRFH}$-$PinX1_{TBM}$ interaction, whereas no measurable interaction was observed between $TRF2_{TRFH}$ and $PinX1_{TBM}$ (fig. S12B). Mutagenesis studies showed that PinX1-L291 and TRF1-*F142* are critical for the interaction, whereas PinX1-P293 is not (fig. S12C). These results are consistent with those of the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction (Fig. 2D) and indicate that PinX1, like TIN2, binds the TRFH domain of TRF1 but not TRF2. Protein sequence database searches showed many instances of telomere-associated proteins containing the F/Y-X-L-X-P motif (fig. S13). Future studies are needed to address whether this motif mediates the TRF1/TRF2 binding of these telomere-associated proteins in vivo.

Our results indicate that binding to the TRFH docking site involves the sequence F/Y-X-L-X-P in shelterin-associated proteins, which contacts the same molecular recognition surface of the TRFH domains of TRF1 and TRF2 with distinct specificities. Because TRF1 and TRF2 play different roles in telomere length homeostasis and telomere protection (*1*), we propose that the TRFH domains of TRF1 and TRF2 function as telomeric protein docking sites that recruit different shelterin-associated factors with distinct functions to the chromosome ends.

References and Notes

1. T. de Lange, *Genes Dev.* **19**, 2100 (2005).
2. M. van Overbeek, T. de Lange, *Curr. Biol.* **16**, 1295 (2006).
3. X. D. Zhu, B. Kuster, M. Mann, J. H. Petrini, T. de Lange, *Nat. Genet.* **25**, 347 (2000).
4. X. D. Zhu *et al.*, *Mol. Cell* **12**, 1489 (2003).
5. S. Smith, I. Giriat, A. Schmitt, T. de Lange, *Science* **282**, 1484 (1998).
6. X. Z. Zhou, K. P. Lu, *Cell* **107**, 347 (2001).
7. T. Nishikawa *et al.*, *Structure* **9**, 1237 (2001).
8. R. Court, L. Chapman, L. Fairall, D. Rhodes, *EMBO Rep.* **6**, 39 (2005).
9. B. Li, S. Oestreich, T. de Lange, *Cell* **101**, 471 (2000).
10. A. Bianchi, S. Smith, L. Chong, P. Elias, T. de Lange, *EMBO J.* **16**, 1785 (1997).
11. L. Fairall, L. Chapman, H. Moss, T. de Lange, D. Rhodes, *Mol. Cell* **8**, 351 (2001).
12. S. H. Kim, P. Kaminker, J. Campisi, *Nat. Genet.* **23**, 405 (1999).
13. C. Lenain *et al.*, *Curr. Biol.* **16**, 1303 (2006).
14. T. H. Lee, K. Perrem, J. W. Harper, K. P. Lu, X. Z. Zhou, *J. Biol. Chem.* **281**, 759 (2006).
15. S. H. Kim *et al.*, *J. Biol. Chem.* **279**, 43799 (2004).
16. J. Z. Ye *et al.*, *J. Biol. Chem.* **279**, 47264 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
19. P. Fotiadou, O. Henegariu, J. B. Sweasy, *Cancer Res.* **64**, 3830 (2004).
20. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with access numbers 3BQO ($TRF1_{TRFH}$-$TIN2_{TBM}$), 3BU8 ($TRF2_{TRFH}$-$TIN2_{TBM}$), and 3BUA ($TRF2_{TRFH}$-$Apollo_{TBM}$).

We thank F. Wang and K. Wan for assistance. Work was supported by an NIH grant (to T.de L.) and an American Cancer Society Research Scholar grant and a Sidney Kimmel Scholar award (to M.L.). Use of Life Sciences Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1151804/DC1
Materials and Methods
SOM Text
Figs. S1 to S14
Table S1
References

16 October 2007; accepted 7 January 2008
Published online 17 January 2008;
10.1126/science.1151804
Include this information when citing this paper.

Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma

Huichen Feng, Masahiro Shuda, Yuan Chang,* Patrick S. Moore*

Merkel cell carcinoma (MCC) is a rare but aggressive human skin cancer that typically affects elderly and immunosuppressed individuals, a feature suggestive of an infectious origin. We studied MCC samples by digital transcriptome subtraction and detected a fusion transcript between a previously undescribed virus T antigen and a human receptor tyrosine phosphatase. Further investigation led to identification and sequence analysis of the 5387–base-pair genome of a previously unknown polyomavirus that we call Merkel cell polyomavirus (MCV or MCPyV). MCV sequences were detected in 8 of 10 (80%) MCC tumors but only 5 of 59 (8%) control tissues from various body sites and 4 of 25 (16%) control skin tissues. In six of eight MCV-positive MCCs, viral DNA was integrated within the tumor genome in a clonal pattern, suggesting that MCV infection and integration preceded clonal expansion of the tumor cells. Thus, MCV may be a contributing factor in the pathogenesis of MCC.

Polyomaviruses have been suspected as potential etiologic agents in human cancer since the discovery of murine polyoma virus (MuPyV) by Gross in 1953 (*1*). However, although polyomavirus infections can produce tumors in animal models, there is no conclusive evidence that they play a role in human cancers (*2*). These small double-stranded DNA viruses [~5200 base pairs (bp)] encode a variably spliced oncoprotein, the tumor (T) antigen (*3*, *4*), and are divided into three genetically distinct groups: (i) avian polyomaviruses, (ii) mammalian viruses related to MuPyV, and (iii) mammalian polyomaviruses related to simian virus 40 (SV40) (*5*). All four known human polyomaviruses [BK virus (BKV), JCV, KIV, and WUV (*6*, *7*)] belong to the SV40 subgroup. In animals, integration of polyomavirus DNA into the host genome often precedes tumor formation (*8*).

Merkel cell carcinoma (MCC) is a neuroectodermal tumor arising from mechanoreceptor Merkel cells (Fig. 1A). MCC is rare, but its incidence has tripled over the past 2 decades in the United States to 1500 cases per year (*9*). It is one of the most aggressive forms of skin cancer; about 50% of advanced MCC patients

Molecular Virology Program, University of Pittsburgh Cancer Institute, University of Pittsburgh, 5117 Centre Avenue, Suite 1.8, Pittsburgh, PA 15213, USA.

*These authors contributed equally to this work. To whom correspondence should be addressed. E-mail: yc70@pitt.edu (Y.C.); psm9@pitt.edu (P.S.M.)

live 9 months or less. Gene expression profiling studies indicate that MCC may comprise two or more clinically similar diseases with distinct etiologies (*10*). Like Kaposi's sarcoma (KS), MCC occurs more frequently than expected among immunosuppressed transplant and AIDS patients (*11*). These similarities to KS, an immune-related tumor caused by KS-associated herpesvirus (*12*), raise the possibility that MCC may also have an infectious origin.

To search for viral sequences in MCC, we used digital transcriptome subtraction (DTS), a methodology we developed that can identify foreign transcripts by using human high-throughput cDNA sequencing data (*13*). We generated two cDNA libraries from a total of four anonymized MCC tumors. One library was prepared with the use of mRNA from a single tumor (MCC347), and the other was prepared with mRNA pooled from three tumors (MCC337, 343, and 346) to increase the likelihood of detecting rare viral sequences (table S1).

From these two libraries, we respectively pyrosequenced 216,599 and 179,135 cDNA sequences (~150 to 200 bp). These 395,734 cDNA sequences were trimmed with LUCY stringency equivalent to PHRED scores of 20 or higher (*14*). Copolymers of adenine or thymidine [poly(A) and poly (T), respectively], dust (low-complexity), human repeat, and primer adaptor sequences were then removed, leaving 382,747 sequences to form a high-fidelity (HiFi) data set. Of these, 380,352 (99.4%) aligned to human RefSeq RNA, mitochondrial, assembled chromosomes, or immunoglobulin sequences in National Center for Biotechnology Information (NCBI) databases. Of the remaining 2395 HiFi candidate sequences, one transcript (DTS1) from MCC347 cDNA aligned with high homology to African green monkey (AGM) lymphotropic polyomavirus (LPyV) and to human BK polyomavirus T antigen sequences. A second DTS transcript (DTS2) had no homology to deposited polyomavirus sequences but was subsequently identified by aligning HiFi candidates to the full-length viral genome (see below). These two sequences define a previously unknown human polyomavirus that we call Merkel cell polyomavirus (MCV or MCPyV) because of its close association with MCC.

Rapid amplification of cDNA ends (3′-RACE) extended DTS1 to three different cDNAs (Fig. 1B): One transcript terminated at a poly(A) site in the T antigen sequence, and two cDNAs read through this weak poly(A) site to form different length fusions with intron 1 of the human receptor tyrosine phosphatase type G gene (*PTPRG*) (GenBank:18860897) on chromosome 3p14.2. Viral integration at this site was confirmed by sequencing DNA polymerase chain reaction (PCR) products with the use of a viral primer and a human *PTPRG* primer. The same three RACE products were independently cloned from MCC348, a lymph node metastasis from the MCC347 primary tumor, indicating that this tumor was seeded from a single tumor cell already positive for the T antigen–PTPRG fusion transcript.

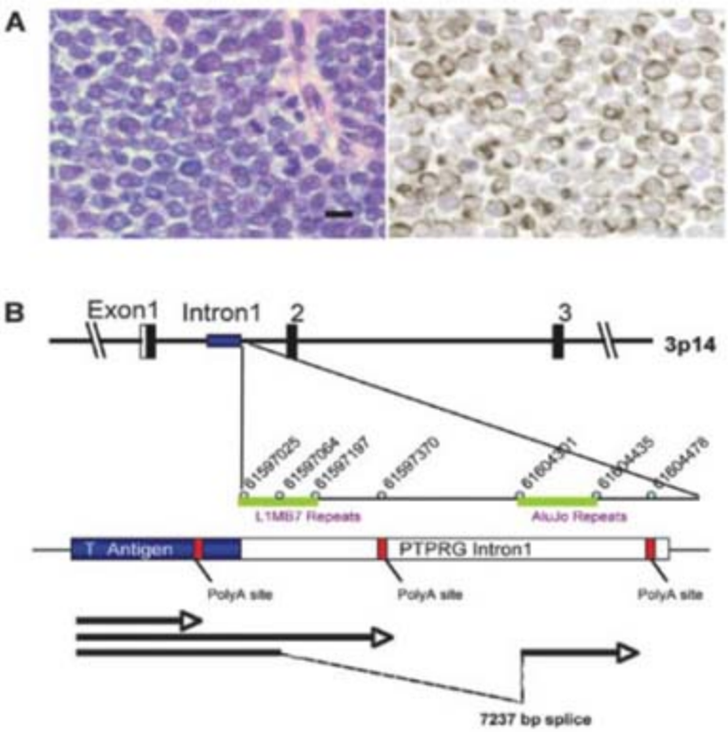


Fig. 1. (**A**) MCC is an aggressive skin cancer derived from Merkel mechanoreceptor cells that expresses neuroendocrine and perinuclear cytokeratin 20 markers, distinguishing it from other small round cell tumors (MCC349, left, hematoxylin and eosin; right, cytokeratin 20 staining, 40×. Scale bar represents 10 μm). (**B**) Discovery of Merkel cell polyomavirus transcripts in (MCC). 3′-RACE mapping of an MCC fusion transcript between the MCV T antigen and human *PTPRG*. A cDNA corresponding to a polyomavirus-like T antigen transcript was found by DTS analysis of MCC. This T antigen cDNA was extended by 3′-RACE to map three mRNA sequences (arrows), one of which terminates at a viral polyadenylation site and two of which extend into flanking human sequence and terminate in intron 1 of the human *PTPRG* gene on chromosome 3p14, indicative of viral DNA integration into the tumor cell genome. The two viral-human chimeric transcripts were generated by read-through of a weak polyadenylation signal in the viral T antigen gene. Identical RACE products were also sequenced from a lymph node metastasis of this primary tumor.

By viral genome walking, we sequenced the complete closed circular genome of MCV (5387 bp, prototype) from tumor MCC350. A second genome, MCV339 (5201 bp), was then sequenced by using MCV-specific primers. The sequences of MCV350 and MCV339 have GenBank accession numbers EU375803 and EU375804, respectively. Both viruses encode sequences with high homology to polyomavirus T antigen, VP1, VP2/3, and replication origin sequences (Fig. 2A). MCV has an early gene expression region [196 to 3080 nucleotides (nt)] containing the T antigen locus with large T and small T open reading frames and a late gene region containing VP1 and VP2/3 open reading frames between 3156 and 5118 nt. The T antigen locus has features conserved with other polyomavirus T antigens, including cr1, DnaJ, pRB1-binding Leu-X-Cys-X-Glu (LXCXE) motif, origin-binding, and helicase/adenosine triphosphatase (ATPase) domains. Mutations in the C terminus of MCV350 and 339 large T open reading frames are predicted to truncate large T protein but are unlikely to affect small T antigen protein expression. The replication origin is highly conserved with that of other polyomaviruses and includes features such as a poly(T) tract and conserved T antigen binding boxes (fig. S1). MCV has highest homology to viruses belonging to the MuPyV subgroup and is most closely related to AGM LPyV (Fig. 2B) (*15*). It is more distantly related to known human polyomaviruses and SV40. The principal differences between MCV350 and MCV339 are a 191-bp (1994 to 2184 nt) deletion in the MCV339 T antigen gene and a 5-bp (5216 to 5220 nt) insertion in the MCV339 late promoter. Excluding these sites, only 41 (0.8%) nucleotides differ between MCV350 and 339.

To investigate the association between MCV infection and MCC, we compared tumors from 10 MCC patients to two tissue control groups. The first control group was composed of unselected tissues from various body sites (including nine skin samples) from 59 patients without MCC (table S2). These samples were taken consecutively on a single surgical day and tested for MCV positivity with two PCR primer sets in the T antigen locus (LT1 and LT3) and one in the VP1 gene (VP1). These primers do not amplify cloned human BKV or JCV genomic DNA or SV40 genome from COS-7 cells. A second control group composed of skin and skin tumor samples from 25 immunocompetent and immunosuppressed patients without MCC were tested with LT1 and VP1 primers (table S2). Samples were randomized and tested in a blinded fashion. Southern blotting of PCR products was performed to increase sensitivity (fig. S2).

Of the 10 MCC tumors from different patients, 8 (80%) were positive for MCV sequences by PCR (Table 1 and table S1). Seven tumors showed robust amplification, and one tumor was positive only after PCR–Southern hybridization. MCC348 (metastasis from MCC347) and

MCC338 (infiltrating tumor from MCC339) were also positive. Two tumors, MCC343 and 346, remained negative after testing with 13 PCR primer pairs spanning the MCV genome. None of the 59 control tissues, including nine skin samples, was positive by PCR alone, but five gastrointestinal tract tissues tested weakly positive after PCR–Southern hybridization (8%, $P < 0.0001$, table S2). Viral T antigen sequences were recovered from three of these samples, confirming low copy number infection. Similarly, only 4 of 25 (16%, $P = 0.0007$, table S2) additional skin and non-MCC skin tumor samples from immunocompetent and immunosuppressed patients tested positive for MCV sequences (Table 2 and table S2).

To determine whether MCV DNA was integrated into the tumor genome, we examined MCC samples by direct Southern blotting without PCR amplification. When MCV DNA in MCC tumor is digested by single-cutter restriction endonucleases, such as EcoRI or BamHI, and probed with viral sequence, four possible patterns are predicted to occur: (i) if the viral DNA exists as freely replicating circular episomes, then a ~5.4 kilobase (kb) band will be present (integrated-concatenated virus will also generate a ~5.4 kb band); (ii) if MCV DNA integrates polyclonally, as might occur during secondary infection of the tumor if MCV is a passenger virus, then diffuse hybridization from different band sizes is expected; (iii) if MCV DNA integrates at one or a few chromosomal sites, then the tumors will have identical or near-identical non-5.4-kb banding patterns; or (iv) if MCV DNA integrates at different chromosomal sites before clonal expansion of the tumor cells, then distinct bands of different sizes will be present (monoclonal viral integration).

Eight of 11 MCC DNA samples (including MCC348 metastasis from MCC347) digested with either BamHI or EcoRI showed robust MCV hybridization, and these corresponded to the same tumors positive by PCR analysis with multiple primers (Fig. 3A and fig. S3). Monoclonal viral integration (pattern iv) was evident with one or both enzymes in six tumors: MCC339, 345, 347, 348, 349, and 352 (solid arrowheads). EcoRI digestion of MCC339, for example, produced two distinct 7.5- and 12.2-kb bands that would arise only if MCV is integrated at a single site in the majority of tumor cells. MCC344 and 350 bands have episomal or integrated-concatemeric bands (open arrowhead, pattern i). MCC352 has a monoclonal integration pattern (solid arrowheads, pattern iv) on BamHI digestion as well as an intense 5.4-kb band (open arrowhead), consistent with an integrated concatemer. All three tumors negative by PCR with ethidium bromide staining (MCC337, 343, and 346) were also negative by direct Southern blotting.

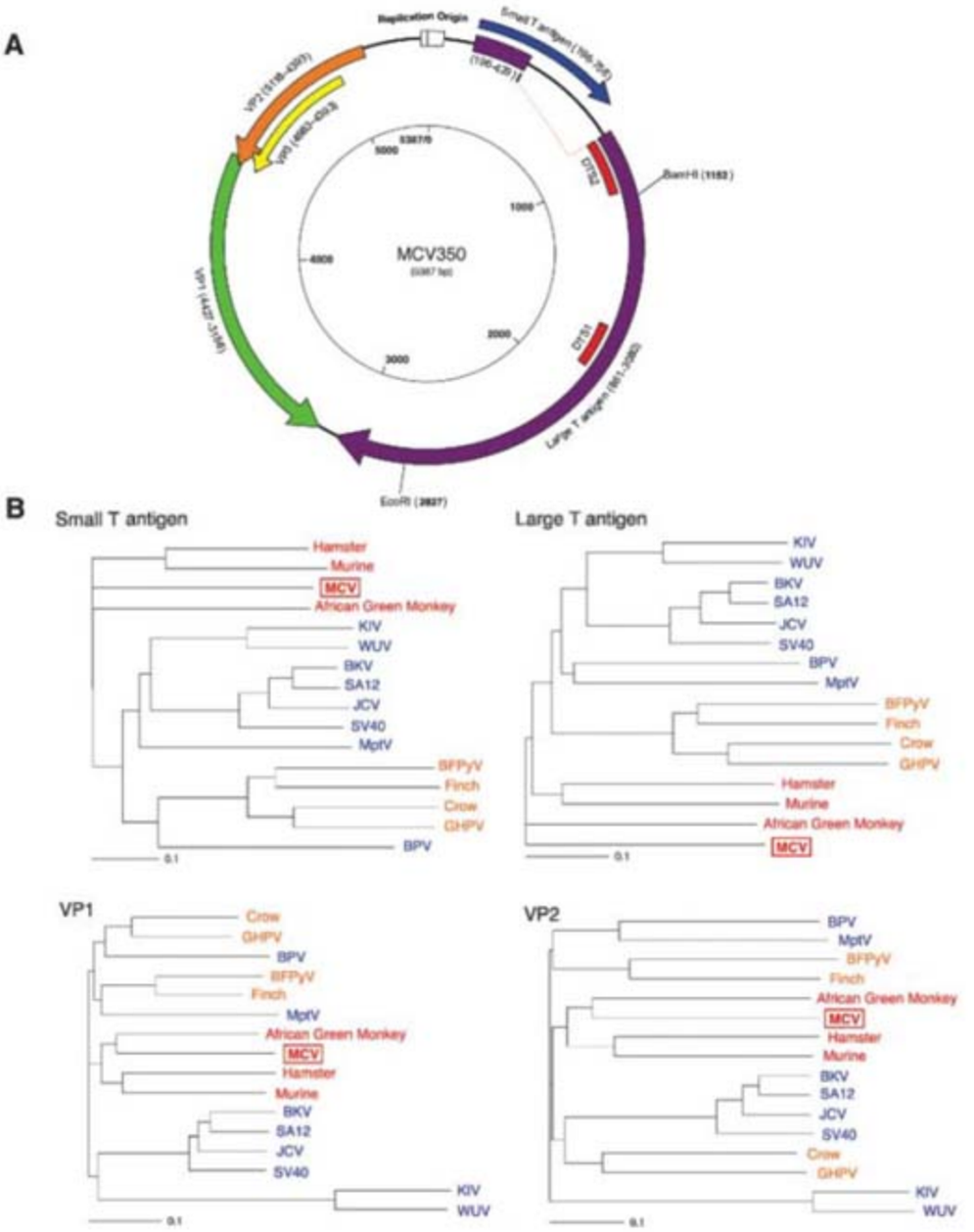


Fig. 2. (**A**) Schematic of MCV genome. Genome walking was used to clone the full MCV genome from tumor MCC350. The genome encodes typical features of a polyomavirus, including large T (purple) and small T (blue) open reading frames. Also shown are predicted VP1 (green) and overlapping VP2 (orange) and VP3 (yellow) genes. DTS1 and DTS2 (red) represent cDNA fragments originally identified by DTS screening. The former was used to identify MCV, and the latter is a spliced transcript with no homology to known polyomavirus sequences. (**B**) Neighbor-joining trees for putative MCV large T, small T, VP1, and VP2 proteins. The four known human polyomaviruses (BKV, JCV, KIV, and WUV) cluster together in the SV40 subgroup (blue), whereas MCV is most closely related to MuPyV subgroup viruses (red). Both subgroups are distinct from the avian polyomavirus subgroup (orange). Scale bars indicate an evolutionary distance of 0.1 amino acid substitutions per position in the sequence.

Table 1. PCR for MCV DNA in MCC tissues. A plus symbol indicates that the sample was strongly positive by ethidium bromide staining only with one or more primers. A minus symbol indicates that the tissue was negative for all primers. Entries with both plus and minus symbols indicate that the sample was negative by ethidium bromide staining but positive after Southern hybridization of PCR products.

MCC cases (n = 10)		
Patient	**Tissue ID**	**MCV positivity**
1	MCC337	−/+
2†	MCC338	+
2	MCC339	+
3	MCC343	−
4	MCC344	+
5	MCC345	+
6	MCC346	−
7	MCC347	+
7‡	MCC348	+
8	MCC349	+
9	MCC350	+
10	MCC352	+
	Total (%)	8/10 (80)

†MCC338 was from an infiltrating tumor in skin tissue adjacent to MCC339 tumor. ‡MCC348 taken from a metastatic lymph node from MCC347.

The Southern blot banding patterns (Fig. 3A) were identical for MCC347 and its metastasis, MCC348, in line with 3′-RACE results (Fig. 1B) and confirming that MCC348 arose as a metastatic clone of MCC347. Because the genomic integration site (the *PTPRG* locus on chromosome 3p14) is mapped for these tumors, we performed Southern blotting with flanking human sequence probes to examine cellular monoclonal integration. NheI-SacI digestion of MCC347 and 348 is predicted to generate a 3.1-kb fragment from the wild-type allele and a 3.9-kb fragment from the allele containing the integrated MCV DNA. Hybridization with a flanking human *PTPRG* sequence probe revealed that the 3.9-kb allele was present in MCC347 and 348 DNA but not in control tissue DNA (Fig. 3B). As predicted, the same fragment hybridized to a MCV T antigen sequence probe, consistent with both cellular and viral monoclonality in this tumor. These results provide evidence that MCV infection and genome integration occurred in this tumor before clonal expansion of tumor cells. MCV in MCC may have some parallels to high-risk human papillomavirus (HPV), which causes cervical cancer mainly after viral episome disruption and integration into the cervical epithelial cell genome (*16*).

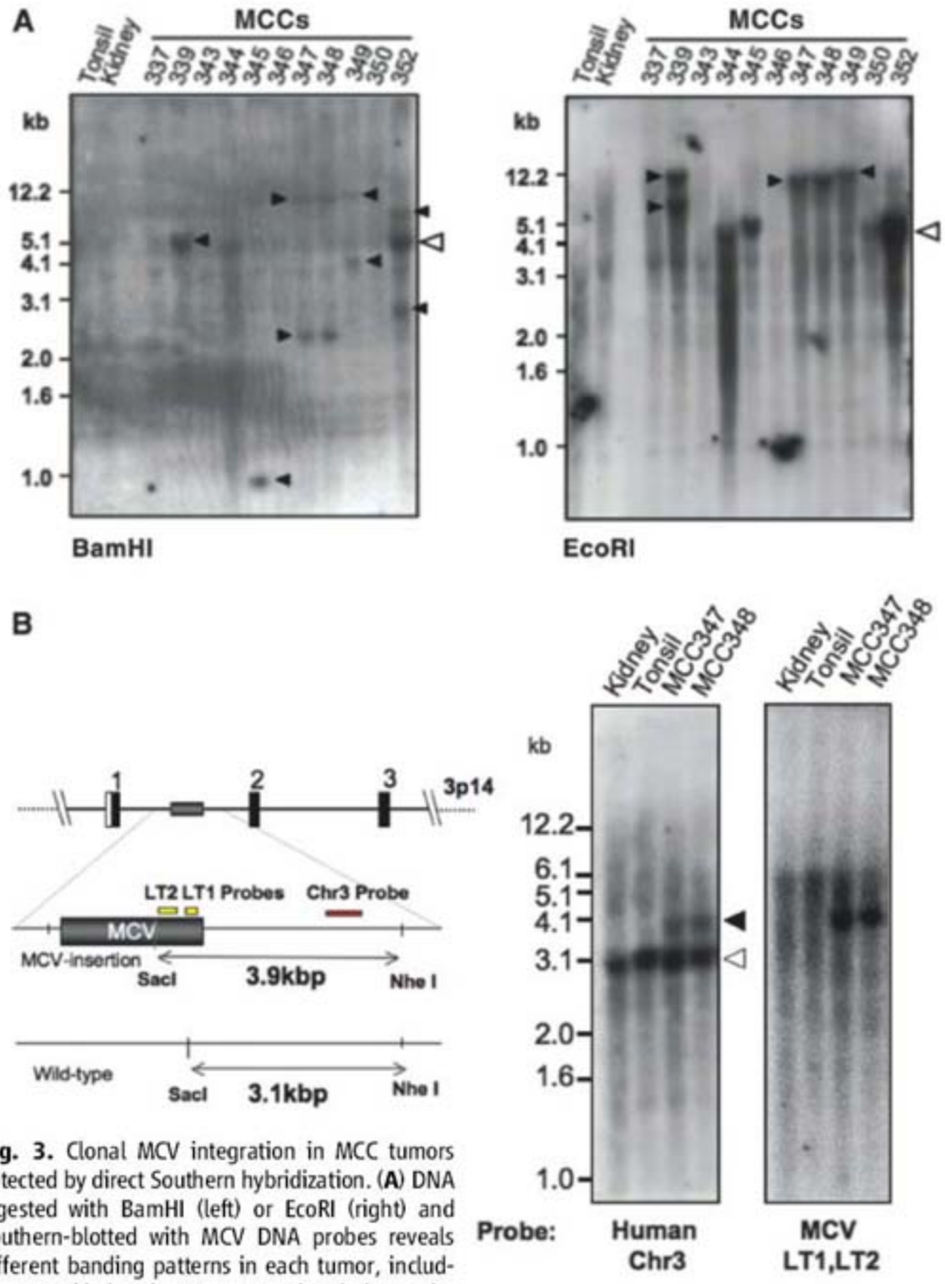


Fig. 3. Clonal MCV integration in MCC tumors detected by direct Southern hybridization. (**A**) DNA digested with BamHI (left) or EcoRI (right) and Southern-blotted with MCV DNA probes reveals different banding patterns in each tumor, including >5.4-kb bands. Open arrowhead shows the expected position for MCV episomal or concatenated-integrated genome (5.4 kb) with corresponding bands present in tumors MCC344 and 350. Tumors MCC339, 345, 347, 348, and 349 have different band sizes and doublet bands (solid arrowheads), consistent with genomic monoclonal integration. MCC352 has a prominent 5.4-kb band as well as higher and lower molecular weight monoclonal integration bands (BamHI), consistent with an integrated concatemer. Tumors MCC337, 343, and 346 have no MCV DNA detected by Southern blotting [bands at 1.5 kb (kidney) and 1.2 kb (MCC346) are artifacts]. (**B**) Viral and cellular monoclonality in MCC347 and 348. Tumor MCC347 and its metastasis MCC348 were digested with SacI and NheI and Southern-blotted with unique human flanking sequence probe [Chr3 (red), left] or viral probes [LT1 and LT2 (yellow), right]. The wild-type human allele is present in all samples at 3.1 kb (left). The MCC tumors, however, have an additional 3.9-kb allelic band formed by MCV DNA insertion into chromosome 3p14. Hybridization with probes for MCV T antigen sequence (yellow, right) generates an identical band.

If MCV plays a causal role in tumorigenesis, it could conceivably do so by several mechanisms, including T antigen expression, insertional mutagenesis, or both. Our DTS results show tumor expression of MCV T antigen, which has conserved DnaJ (*4*), pocket protein-binding LXCXE (*17*), and pp2A-binding (*18*, *19*) domains previously shown to play roles in polyomavirus-induced cell transformation. Mutational disruption of the *PTPRG* gene, which is suspected to be a tumor suppressor (*20*), could also play a role in MCC, although our Southern blot data suggest that MCV integration occurs at various genomic sites in different MCC tumors.

Our study validates the utility of DTS for the discovery of cryptic human viruses, but it has also revealed some limitations of the approach. Of the four tumors we sampled, only one (MCC347) was infected at high copy number. MCV transcripts in this tumor were present at 10 transcripts per million or about 5 transcripts per tumor cell. In future searches for other directly transforming tumor viruses (*21*), DTS should be used on multiple highly uniform samples sequenced to a depth of 200,000 transcripts or greater. Because DTS is quantitative, it is less likely to be useful in its current form for discovery of low-abundance viruses in autoimmune disorders or other chronic infectious diseases. Discovery of MCV by DTS nonetheless shows that DTS and related approaches (*22*) are promising methods to identify previously unknown human tumor viruses.

References and Notes

1. L. Gross, *Proc. Soc. Exp. Biol. Med.* **83**, 414 (1953).
2. D. L. Poulin, J. A. DeCaprio, *J. Clin. Oncol.* **24**, 4356 (2006).
3. S. M. Dilworth, *Nat. Rev. Cancer* **2**, 951 (2002).
4. J. M. Pipas, *J. Virol.* **66**, 3979 (1992).
5. K. A. Crandall, M. Perez-Losada, R. G. Christensen, D. A. McClellan, R. P. Viscidi, *Adv. Exp. Med. Biol.* **577**, 46 (2006).
6. T. Allander *et al.*, *J. Virol.* **81**, 4130 (2007).
7. A. M. Gaynor *et al.*, *PLoS Pathog.* **3**, e64 (2007).
8. D. Hollanderova, H. Raslova, D. Blangy, J. Forstova, M. Berebbi, *Int. J. Oncol.* **23**, 333 (2003).
9. B. Lemos, P. Nghiem, *J. Investig. Dermatol.* **127**, 2100 (2007).
10. M. Van Gele *et al.*, *Oncogene* **23**, 2732 (2004).
11. E. A. Engels, M. Frisch, J. J. Goedert, R. J. Biggar, R. W. Miller, *Lancet* **359**, 497 (2002).
12. Y. Chang *et al.*, *Science* **266**, 1865 (1994).
13. H. Feng *et al.*, *J. Virol.* **81**, 11332 (2007).
14. H. H. Chou, M. H. Holmes, *Bioinformatics* **17**, 1093 (2001).
15. M. Pawlita, A. Clad, H. zur Hausen, *Virology* **143**, 196 (1985).
16. M. Durst, L. Gissmann, H. Ikenberg, H. zur Hausen, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3812 (1983).
17. J. A. DeCaprio *et al.*, *Cell* **54**, 275 (1988).
18. D. C. Pallas *et al.*, *Cell* **60**, 167 (1990).
19. G. Walter, R. Ruediger, C. Slaughter, M. Mumby, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2521 (1990).
20. D. M. Pitterle, E. M. Jolicoeur, G. Bepler, *In Vivo (Athens)* **12**, 643 (1998).

Table 2. PCR for MCV DNA in comparison control tissues (n = 84). For detailed description of tissues and tissue sites, see table S2. MCV positivities marked with plus and minus symbols together are as in Table 1. For the various body site tissues, there were 59 samples; for the skin and skin tumor tissues, the sample size was 25 (table S2).

	MCV positivity
Various body site tissues	
Total MCV negative (%)	54/59 (92)
Total MCV positive (%)	5/59 (8)
Appendix control 1	−/+
Appendix control 2	−/+
Gall bladder	−/+
Bowel	−/+
Hemorrhoid	−/+
Skin and skin tumor tissues	
Total MCV negative (%)	21/25 (84)
Total MCV positive (%)	4/25 (16)
Skin	−/+
KS skin tumor 1	−/+
KS skin tumor 2	−/+
KS skin tumor 3	−/+

21. J. Parsonnet, in *Microbes and Malignancy*, J. Parsonnet, Ed. (Oxford Univ. Press, New York, 1999), pp. 3–18.
22. Y. Xu *et al.*, *Genomics* **81**, 329 (2003).

23. We thank the National Cancer Institute–supported Cooperative Human Tissue Network for tissues used in this study, M. Aquafondata for tissue staining, P. S. Schnable for sharing cDNA data sets used in DTS pilot testing, O. Gjoerup and R. D. Wood for helpful comments, and J. Zawinul for help with the manuscript. Supported in part by funds from NIH R33CA120726 and the Pennsylvania Department of Health. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1152586/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S5
References

5 November 2007; accepted 8 January 2008
Published online 17 January 2008;
10.1126/science.1152586
Include this information when citing this paper.

Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation

Jun Z. Li,[1,2]*† Devin M. Absher,[1,2]* Hua Tang,[1] Audrey M. Southwick,[1,2] Amanda M. Casto,[1] Sohini Ramachandran,[4] Howard M. Cann,[5] Gregory S. Barsh,[1,3] Marcus Feldman,[4]‡ Luigi L. Cavalli-Sforza,[1]‡ Richard M. Myers[1,2]‡

Human genetic diversity is shaped by both demographic and biological factors and has fundamental implications for understanding the genetic basis of diseases. We studied 938 unrelated individuals from 51 populations of the Human Genome Diversity Panel at 650,000 common single-nucleotide polymorphism loci. Individual ancestry and population substructure were detectable with very high resolution. The relationship between haplotype heterozygosity and geography was consistent with the hypothesis of a serial founder effect with a single origin in sub-Saharan Africa. In addition, we observed a pattern of ancestral allele frequency distributions that reflects variation in population dynamics among geographic regions. This data set allows the most comprehensive characterization to date of human genetic variation.

In the past 30 years, the ability to study DNA sequence variation has dramatically increased our knowledge of the relationships among and history of human populations. Analyses of mitochondrial, Y chromosomal, and autosomal markers have revealed geographical structuring of human populations at the continental level (*1*–*3*) and suggest that a small group of individuals migrated out of eastern Africa and their descendants subsequently expanded into most of today's populations (*3*–*6*). Despite this progress, these studies were limited to a small fraction of the genome, to limited populations, or both, and yield an incomplete picture of the relative importance of mutation, recombination, migration, demography, selection, and random drift (*7*–*10*). To substantially increase the genomic and population coverage of past studies (e.g., the HapMap Project), we have examined more than 650,000 single-nucleotide polymorphisms (SNPs) in samples from the Human Genome Diversity Panel (HGDP-CEPH), which represents 1064 fully consenting individuals from 51 populations from sub-Saharan Africa, North Africa, Europe, the Middle East, South/Central Asia, East Asia, Oceania, and the Americas (*11*). This data set is freely available (*12*) and allows a detailed characterization of worldwide genetic variation.

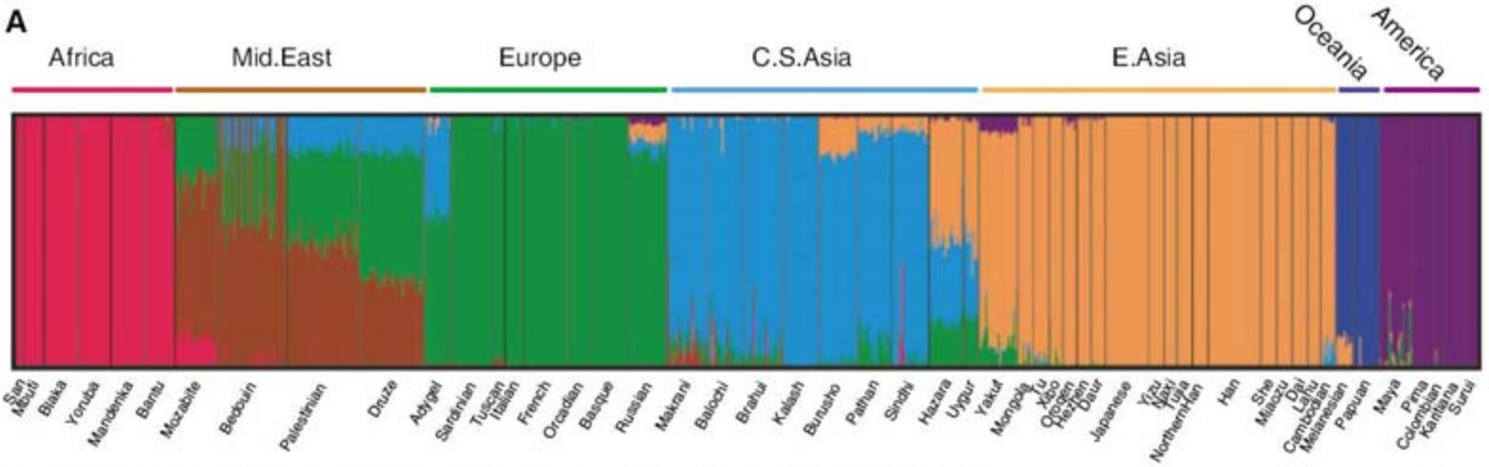
We first studied genetic ancestry of each individual without using his/her population identity. This analysis considers each person's genome as having originated from K ancestral but unobserved populations whose contributions are described by K coefficients that sum to 1 for each individual. To increase computational efficiency, we developed new software, *frappe*, that implements a maximum likelihood method (*13*) to analyze all 642,690 autosomal SNPs in 938 unrelated and successfully genotyped HGDP-CEPH individuals (*14*). Figure 1A shows the results for K = 7; those for K = 2 through 6 are in fig. S1. At K = 5, the 938 individuals segregate into five continental groups, similar to those re-

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [2]Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [3]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [4]Department of Biological Sciences, Stanford University, Stanford, CA 94305–5120, USA. [5]Foundation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 75010 Paris, France.

*These authors contributed equally to this work.
†Present address: Department of Human Genetics, University of Michigan, 5789A MS II, Ann Arbor, MI 48109–5618, USA.
‡To whom correspondence should be addressed. E-mail: marc@charles.stanford.edu (M.F.); cavalli@stanford.edu (L.L.C.S.); myers@shgc.stanford.edu (R.M.M.)

ported in a microsatellite-based study of the same panel (*3*). At $K = 6$, the new component accounts for a major portion of ancestry for individuals from South/Central Asia, separating this region from the Middle East and Europe. This result differs from that in (*3*), where the sixth component contained the Kalash individuals, but South/Central Asia, the Middle East, and Europe were not clearly distinguished unless analyzed separately from the rest of the world. At $K = 7$, the new component occurs at highest proportions in the Middle Eastern populations, separating them from European populations. In many populations, ancestry is derived predominantly from one of the inferred components, whereas in others, especially those in the Middle East and South/Central Asia, there are multiple sources of ancestry. For example, Palestinians, Druze, and Bedouins have contributions from the Middle East, Europe, and South/Central Asia. Burusho, Pathan, and Sindhi have an East Asian contribution. Hazara and Uygur share a similar profile of combined South/Central Asian, East Asian, and European ancestry. In East Asia, only the Yakuts share ancestry with both Europe and America, although these contributions are small. Although much of sub-Saharan Africa, Europe, and East Asia appears to be homogeneous in Fig. 1A, finer substructures can be detected when individual regions are analyzed separately. For example, we identified two components that separate the 16 East Asian populations and correspond to a north-south genetic gradient (fig. S2A). Han Chinese can be divided into a southern and a northern group. A similar analysis for South/Central Asia is shown in fig. S2B.

Mixed ancestries inferred from genetic data can often be interpreted as arising from recent admixture among multiple founder populations. In the current setting, however, the estimated mixed ancestry can be due either to recent admixture or to shared ancestry before the diver-



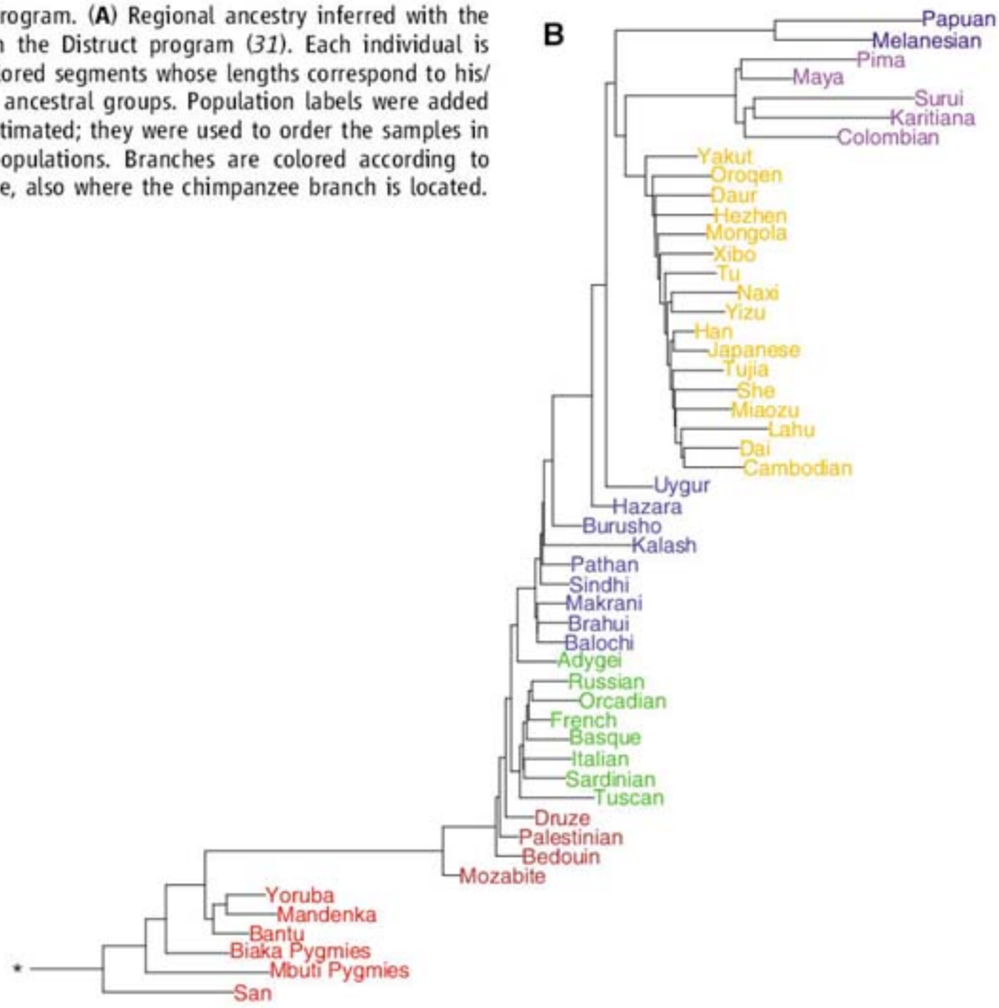


Fig. 1. Individual ancestry and population dendrogram. (**A**) Regional ancestry inferred with the *frappe* program at $K = 7$ (*13*) and plotted with the Distruct program (*31*). Each individual is represented by a vertical line partitioned into colored segments whose lengths correspond to his/her ancestry coefficients in up to seven inferred ancestral groups. Population labels were added only after each individual's ancestry had been estimated; they were used to order the samples in plotting. (**B**) Maximum likelihood tree of 51 populations. Branches are colored according to continents/regions. * indicates the root of the tree, also where the chimpanzee branch is located.

gence of two populations but without subsequent gene flow between them. For example, the European and Asian ancestries seen in Uygur and Hazara populations are likely due to relatively recent admixture, whereas the inferred Native American ancestry in Yakuts and Russians likely reflects shared ancestry before the predecessors of the Native Americans crossed the Bering Strait. The Middle Eastern populations may have experienced both continuous gene flow and shared ancestry with the rest of Eurasia.

Because individuals belonging to the same recognized population almost always show similar ancestry proportions (Fig. 1A and fig. S2), it is meaningful to statistically evaluate the genetic relationships among populations. We calculated the Wright's fixation indices Fsts among the 51 populations from the population allele frequencies across all autosomal SNPs (*15*) and constructed a phylogenetic tree by the maximum likelihood method (*16*), using orthologous chimpanzee alleles as the outgroup. The sub-Saharan African populations are located nearest to the root of the tree (Fig. 1B), outward from which are branches that correspond, sequentially, to populations from North Africa, the Middle East, Europe, South/Central Asia, Oceania, America, and East Asia. This population tree shows not only major splits between different continents but also sublineages within continents (*14*) consistent with the ancestry analysis shown above as well as with results from microsatellite markers (*17*). The branching pattern largely agrees with the approximate order of human expansion (*2*) and supports the "out of Africa" model of human origin.

We performed principal component analyses (PCA) on the Fst matrix to capture a major portion of genetic variability. The first and second PCs explain 59% and 26% of the Fst variation, respectively (fig. S3A) and separate the 51 populations into the known continental groups, with the first PC primarily describing the contrast between sub-Saharan Africans and non-Africans and the second driven by the East-West difference in Eurasia. The third and fourth PCs distinguish the Native American and Oceanian populations, respectively (figs. S3, C and D). The regional clusters are more clearly separated than was possible with 782 microsatellites (*16*). A PCA plot of the 938 unrelated individuals (fig. S3B) is similar to the 51-population plot and illustrates the regional clusteredness at the individual level.

The PCA for individual continents/regions clearly delineates fine-scale population structure. In Fig. 2A, the eight European populations, including the central populations (Orcadian, French, Northern Italian from Bergamo, and Tuscan) which were previously indistinguishable with fewer markers, can be separated (*3*). In Fig. 2B, the four populations from the Middle East are distinguished; the Bedouins can be divided into two subgroups, one of which is similar to the Palestinians. The PC1-PC2 plots for four other continental groups and descriptions and interpretations are in (*14*) and figs. S4 and S5. These individual-level results, along with ancestry analyses in Fig. 1A and fig. S2, indicate that although some populations have limited sample size (<10), the population structuring appears robust.

We carried out an analysis of molecular variance (AMOVA) (*18*, *19*) to partition overall genetic variation into three components: within-population (WP), among-population-within-group (i.e., geographical region) (AP/WG), and among geographical region (AG). The 51 populations are assigned to the seven geographical regions shown in Fig. 1A. The results are similar among autosomal chromosomes: the WP, AP/WG, and AG components explain 88.9 ± 0.3%, 2.1 ± 0.05%, and 9.0 ± 0.3% (mean ± SD across 22 chromosomes) of the variance, respectively (Fig. 3A). For comparison, the WP, AP/WG, and AG components for 783 microsatellite markers are 94.0%, 2.3%, and 3.7%, respectively (*3*, *5*). The difference between the SNP-based estimates (this study) and the microsatellite-based results can be partly explained by higher mutation rates of microsatellites, which are more driven by shorter-term evolutionary processes. For X chromosome (ChrX) SNPs, the WP, AP/WG, and AG components are 84.7%, 2.4%, and 12.9% (Fig. 3A), consistent with estimates based on ChrX microsatellites (*20*). The greater AG component for ChrX than autosomes is discussed in (*14*) and fig. S6. Together, these results reaffirm that within-population variation accounts for most of the genetic diversity in humans. However, the between-population variance is sufficient to reveal consistent population structure because subtle but nonrandom differences between populations accumulate over a large number of loci and yield principal components that can account for a major portion of the variation (*21*).

We compared SNP haplotype heterozygosity across populations and found, consistent with earlier reports (*22*), that it is highest in sub-Saharan Africa and decreases steadily with distance from this region (Fig. 3B). The mean heterozygosity across autosomal haplotypes (using 295 haplotype blocks in Chr16) (*14*) is negatively correlated with distance from Addis Ababa, Ethiopia (*5*, *23*), with a correlation coefficient r of –0.91 and a slope of -1.1×10^{-5} per km (Fig. 3B). This trend is consistent with a serial founder effect, a scenario in which population

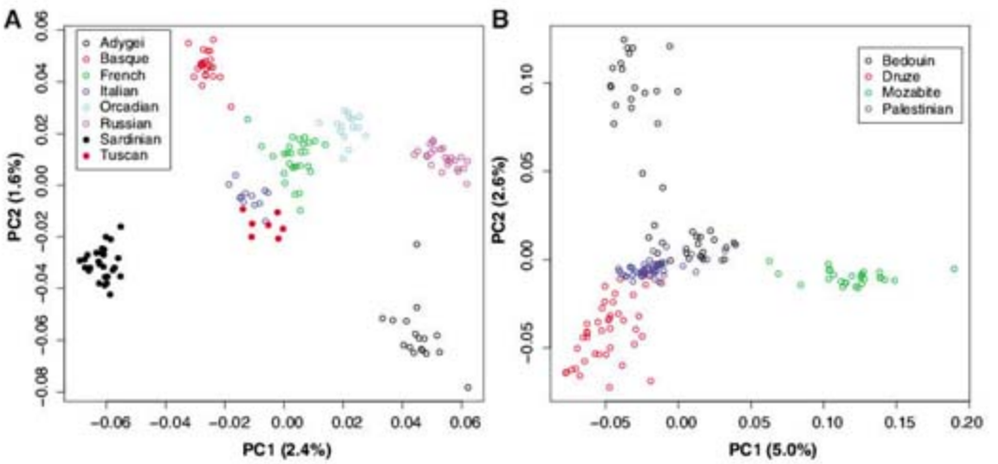


Fig. 2. Fine-scale population structure principal component analyses in two geographic regions, using all autosomal SNPs. (**A**) Europe. (**B**) The Middle East.

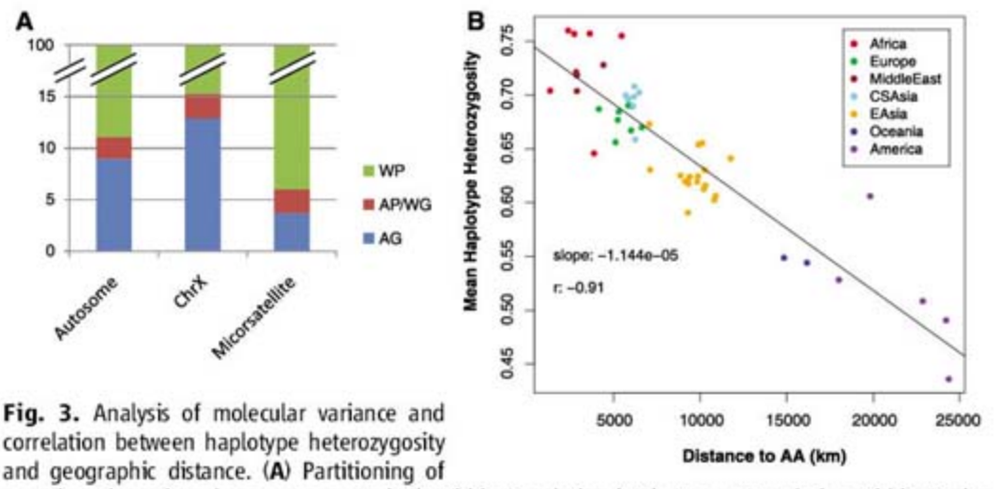


Fig. 3. Analysis of molecular variance and correlation between haplotype heterozygosity and geographic distance. (**A**) Partitioning of genetic variance into three components (*18*): Within-Population (WP), Among-Population-Within-Region (AP/WG), and Among-Region (AG), by using autosomal SNPs, microsatellite markers, and ChrX SNPs, respectively. (**B**) SNP haplotype heterozygosity versus geographic distances from Addis Ababa (AA), Ethiopia. The linear regression slope is indicated along with the Pearson correlation r.

expansion involves successive migration of a small fraction of individuals out of the previous location, starting from a single origin in sub-Saharan Africa. For ChrX haplotypes (using 453 haplotype blocks), the correlation and slope are −0.85 and -1.3×10^{-5} per km, respectively; the slightly higher geographic gradient (i.e., steeper slope) than for autosomes agrees with the higher ChrX Fsts (fig. S6). These values are similar to those reported for microsatellites (*5*): $r = -0.87$, slope = -6.52×10^{-6} per km, and to SNP-based heterozygosities: r = −0.81, slope = -3.8×10^{-6} (r = −0.93 when only non-African populations are considered). SNP-based heterozygosities depend on allele frequencies and are affected by ascertainment bias, whereas the haplotype and microsatellite heterozygosities are less affected as a result of their greater polymorphism (*22*).

By genotyping two chimpanzee samples, we were able to define the putative ancestral allele for ~95.5% of the SNPs in the 650 K panel. We compared the distribution of these ancestral allele frequencies (AAFs) among the 51 populations. Figure 4A shows four examples of the AAF spectrum for Yoruba, French, Chinese, and Japanese populations. Yorubans and other sub-Saharans have more SNPs with high AAFs (>0.6, on the right of the distribution) and fewer with low AAFs, producing a steeper slope of SNP counts in the midrange of AAF spectrum. The slopes within 20 to 80% AAF are plotted in Fig. 4B for all 51 populations, showing a progressive reduction moving away from Africa, from ~0.04 in sub-Saharan Africa, to ~0.03 in Eurasia, ~0.02 in East Asia, and ~0.01 in Oceania and the Americas. This steady flattening of the AAF distribution may be related to the SNP panel used (which primarily includes common SNPs in Europe, East Asia, and sub-Saharan Africa), but the ascertainment scheme alone cannot explain the entire trend, as the SNPs analyzed by the International HapMap Project show a similar phenomenon (*24*). In particular, the AAF spectra of ENCODE (ENCyclopedia of DNA Elements) regions, where genotyped SNPs were discovered by resequencing, follow a similar pattern (fig. S7), where the 20 to 80% slopes in HapMap Chinese and Japanese populations are about half of that in the Yoruban population.

The flattening of the AAF spectrum reflects the interplay of multiple demographic forces and may yield clues to the history of individual populations. Generally, populations that had a small effective population size and/or experienced a severe bottleneck would have more pronounced genetic drift, resulting in a more rapid increase in derived allele frequencies. Populations that maintained a large size or experienced expansion would tend to preserve the ancestral states of the variant loci. Theoretical work (*25*), empirical data, and simulation (*26*) have shown that demography plays a major role in the change of the AAF spectrum over time. Our result is consistent with the serial founder model, in which non-African populations form a sequential chain of colonies. Those that are more peripheral and younger have relatively smaller effective sizes, and perhaps experienced greater selective pressure. For example, the European and Asian spectra can be explained by a reduction of population size followed by recent recovery (a bottleneck), whereas the spectrum for an African-American population suggests a history of moderate but uninterrupted expansion (*26*, *27*).

Compared to the HapMap panel, HGDP-CEPH includes Oceanian and American populations, as well as a dense collection from the Middle East and South/Central Asia. Characterization of the added populations is important for studying evolution and disease processes not only in these populations but also in those that share common ancestry with them due to recent migration (e.g., U.S. Latino populations). HGDP-CEPH is not a random sample of the world's populations: Some parts of the world (e.g., China and Pakistan) are more densely covered than others (Africa, the Americas, and Oceania). Many populations have been isolated from each other by geography or custom. The observation that they are genetically distinguishable suggests that

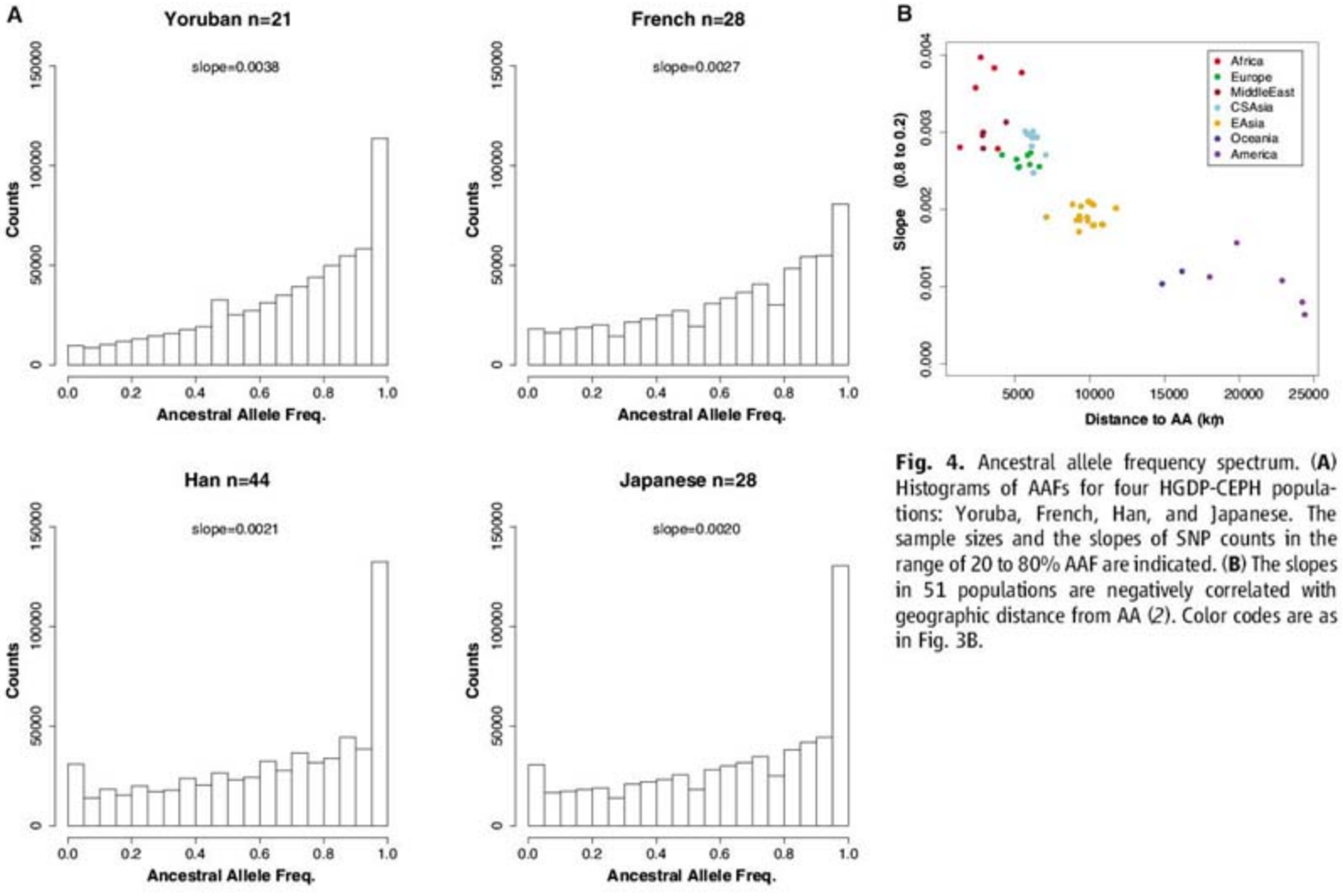


Fig. 4. Ancestral allele frequency spectrum. (**A**) Histograms of AAFs for four HGDP-CEPH populations: Yoruba, French, Han, and Japanese. The sample sizes and the slopes of SNP counts in the range of 20 to 80% AAF are indicated. (**B**) The slopes in 51 populations are negatively correlated with geographic distance from AA (*2*). Color codes are as in Fig. 3B.

self-reported ancestry is sufficiently accurate for assessing population stratification in disease studies, except for those involving recent admixture (*3*, *28*). These results, however, say nothing about the origin and distribution of human phenotypic variation. The observed population structure can be largely explained by random drift at neutral loci. Nevertheless, some regions of the genome may have experienced accelerated divergence due to local selection (*9*, *24*, *29*) as anatomically modern humans spread around the globe during the past 100,000 years, adapting to a wide range of habitats and climates. The population richness of HGDP-CEPH makes it possible to detect correlation between genomic variation and local environmental and/or phenotypic variation (*30*), thus leading to more detailed understandings of selective forces acting in different regions of the world.

References and Notes

1. M. Bamshad *et al.*, *Nat. Rev. Genet.* **5**, 598 (2004).
2. L. L. Cavalli-Sforza *et al.*, *Nat. Genet.* **33** (suppl.), 266 (2003).
3. N. A. Rosenberg *et al.*, *Science* **298**, 2381 (2002).
4. A. M. Bowcock *et al.*, *Nature* **368**, 455 (1994).
5. S. Ramachandran *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15942 (2005).
6. N. A. Rosenberg *et al.*, *PLoS Genet.* **1**, e70 (2005).
7. C. D. Bustamante *et al.*, *Nature* **437**, 1153 (2005).
8. L. L. Cavalli-Sforza, *Annu. Rev. Genomics Hum. Genet.* **8**, 1 (2007).
9. B. F. Voight *et al.*, *PLoS Biol.* **4**, e72 (2006).
10. S. H. Williamson *et al.*, *PLoS Genet.* **3**, e90 (2007).
11. H. M. Cann *et al.*, *Science* **296**, 261 (2002).
12. Human Genome Diversity Panel, http://shgc.stanford.edu/hgdp or ftp://ftp.cephb.fr/hgdp_supp1.
13. H. Tang *et al.*, *Genet. Epidemiol.* **28**, 289 (2005).
14. Materals and methods are available as supporting material on *Science* Online.
15. J. Reynolds *et al.*, *Genetics* **105**, 767 (1983).
16. J. Felsenstein, *Am. J. Hum. Genet.* **25**, 471 (1973).
17. L. A. Zhivotovsky *et al.*, *Am. J. Hum. Genet.* **72**, 1171 (2003).
18. L. Excoffier *et al.*, *Genetics* **131**, 479 (1992).
19. B. Weir, *Genetic Data Analysis II* (Sinauer, Sunderland, MA, 1996).
20. S. Ramachandran *et al.*, *Hum. Genomics* **1**, 87 (2004).
21. A. W. Edwards, *Bioessays* **25**, 798 (2003).
22. D. F. Conrad *et al.*, *Nat. Genet.* **38**, 1251 (2006).
23. F. Prugnolle *et al.*, *Curr. Biol.* **15**, R159 (2005).
24. International HapMap Consortium, *Nature* **437**, 1299 (2005).
25. M. Kimura *et al.*, *Genetics* **75**, 199 (1973).
26. G. T. Marth *et al.*, *Genetics* **166**, 351 (2004).
27. A. Keinan *et al.*, *Nat. Genet.* **39**, 1251 (2007).
28. H. Tang *et al.*, *Am. J. Hum. Genet.* **76**, 268 (2005).
29. S. A. Tishkoff *et al.*, *Nat. Genet.* **39**, 31 (2007).
30. A. Manica *et al.*, *Nature* **448**, 346 (2007).
31. N. A. Rosenberg, *Mol. Ecol. Notes* **4**, 137 (2004).

32. We thank H. Greely for support and helpful discussions, A. Aggarwal and S. Brady for technical assistance, and the Biological Resource Center of the Foundation Jean Dausset-CEPH for preparing and formatting HGDP-CEPH diversity panel DNAs. The BeadArrays used for genotyping the HGDP-CEPH samples were provided by 23andMe and Illumina, Inc. All authors declare no competing financial interest. M.W.F. and H.T. were supported in part by NIH grants GM28016 and GM073059, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1100/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Table S1
References

3 December 2007; accepted 17 January 2008
10.1126/science.1153717

Spine-Type–Specific Recruitment of Newly Synthesized AMPA Receptors with Learning

Naoki Matsuo,*† Leon Reijmers, Mark Mayford*

The stabilization of long-term memories requires de novo protein synthesis. How can proteins, synthesized in the soma, act on specific synapses that participate in a given memory? We studied the dynamics of newly synthesized AMPA-type glutamate receptors (AMPARs) induced with learning using transgenic mice expressing the GluR1 subunit fused to green fluorescent protein (GFP-GluR1) under control of the *c-fos* promoter. We found learning-associated recruitment of newly synthesized GFP-GluR1 selectively to mushroom-type spines in adult hippocampal CA1 neurons 24 hours after fear conditioning. Our results are consistent with a "synaptic tagging" model to allow activated synapses to subsequently capture newly synthesized receptor and also demonstrate a critical functional distinction in the mushroom spines with learning.

AMPA-type glutamate receptors (AMPARs) are the primary mediators of fast excitatory transmission in the mammalian brain and play a key role in long-term potentiation (LTP) (*1*). There is experience-dependent synaptic trafficking of GluR1 homomers in barrel cortex and lateral amygdala of juvenile rats (12 to 25 days old) (*2*, *3*). These results suggest that changes in the strength of excitatory synapses contribute to learning, possibly through rapid synaptic insertion of preexisting GluR1-containing AMPARs, although it is unclear whether similar mechanisms are present in the adult (*4*–*6*). Short-term memories are lost if new protein synthesis is inhibited, indicating a requirement for de novo protein synthesis to consolidate the memories into a long-lasting form (*7*). This has raised a fundamental question: How do new proteins, synthesized in the soma, exert their effect on specific synapses involved in synaptic or behavioral plasticity? It has been suggested that stimulation produces a "tag" at the activated synapses to allow the capture of newly synthesized plasticity-related proteins at later time points to maintain the increased synaptic strength (*8*). Synthesis of AMPARs is increased after LTP induction (*9*), suggesting a possible role in this process. To determine whether these mechanisms might be operating in the adult brain in vivo, we developed transgenic mice to monitor the trafficking and turnover of newly synthesized AMPARs induced at the time of learning in a fear conditioning paradigm.

The transgenic mice express a GluR1 subunit fused to green fluorescent protein (GFP-GluR1) in a doxycycline (Dox)–regulated (*10*) and neuronal activity–dependent manner and are referred to as GFP-GluR1$^{c\text{-}fos}$ Tg mice (Fig. 1A). We used the activity-dependent *c-fos* promoter (*11*, *12*) to induce a rapid and transient expression of GFP-GluR1 in response to environmental stimuli and to focus on the molecular and cellular events specifically in those neurons activated by the behaviorally relevant events. Fear conditioning (*13*) was used in adult animals to evaluate regulated expression in the dorsal hippocampus. GFP fluorescence, immunohistochemical, and immunoblot analysis revealed that GFP-GluR1 expression was negligible in the presence of Dox, even in fear-conditioned mice (Fig. 1, B, C, F, and G), showing successful transgene suppression by Dox. The GFP expression seen in some of the cell nuclei is derived from a cfos-nlsGFP transgene, which is expressed independently of the Dox-regulated system. To test the inducibility of GFP-GluR1$^{c\text{-}fos}$ Tg by behavioral training, mice were removed from Dox for 4 days and were either fear conditioned or allowed to remain in the homecage. At 24 hours after conditioning, prominent GFP-GluR1 expression was detected in the fear-conditioned animals relative to homecage controls (Fig. 1D-G), showing activity-dependent expression of GFP-GluR1 in the absence of Dox. The induced somatic

Department of Cell Biology and Institute for Childhood and Neglected Diseases, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: n-matsuo@fujita-hu.ac.jp (N.M.); mmayford@scripps.edu (M.M.)

†Present address: Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan.

GFP-GluR1 was detected in ~25% of CA1 pyramidal neurons within 2 hours of fear conditioning and was maintained for more than 24 hours (Fig. 1F and fig. S1). One concern is the potential for overexpressed receptors to function aberrantly. In many previous studies of recombinant GluR1 trafficking, the receptor formed exclusively homo-oligomers (*2*, *3*, *14*, *15*), whereas the endogenous receptor complex is primarily a heteromer (*16*). The induced expression of the GFP-GluR1 was ~3% of endogenous GluR1 levels (Fig. 1G). GFP-GluR1 was coimmunoprecipitated with the endogenous GluR2 subunit (Fig. 1H), demonstrating the formation of the natural heteromeric forms. The dendritic transport rate and half-life of GFP-GluR1 were consistent with that determined previously for endogenous receptor in vitro (*17*–*19*) (figs. S1 and S2).

In contextual fear conditioning, animals learn the association between a specific environment (context) and an aversive stimulus such as a footshock. Brief training produces a stable and long-lasting memory that requires the hippocampus (*13*). We used fear conditioning in the GFP-GluR1$^{c\text{-}fos}$ Tg mice to test for learning-associated synaptic trafficking of newly synthesized AMPARs in the adult hippocampus. Mice were removed from Dox and were fear conditioned to elicit a contextual long-term memory and to induce synthesis of GFP-GluR1 in the activated neurons (FC24h) (Fig. 2A). Control mice received either context exposure without footshocks (CT24h) (Fig. 2A) or unpaired context-shock training, which failed to produce a contextual fear memory (UP24h) (Fig. 2A and fig. S3). Mice were returned to their homecage and treated with high Dox (6 g/kg of mouse chow) to rapidly suppress further expression of GFP-GluR1. Twenty-four hours after training, brain slices were fixed and stained with antibody to GFP, without permeabilization, to detect surface GFP-GluR1 expression (fig. S4). A subset of neurons in the CA1 pyramidal cell layer of the dorsal hippocampus was stained with the lipophilic fluorescent dye, DiI, to visualize the entire dendritic arbor and spine distribution on individual neurons (Fig. 2B-D). Consistent with previous studies examining endogenous c-Fos induction (*20*), all groups of mice showed a similar number of GFP-GluR1^{+} CA1 neurons (Fig. 1F). Levels of GFP-GluR1 expression were also similar in the fear-conditioned and context-exposure group (Fig. 2E). This design allows us to induce the receptor at the time of learning and to follow its distribution to synapses at later

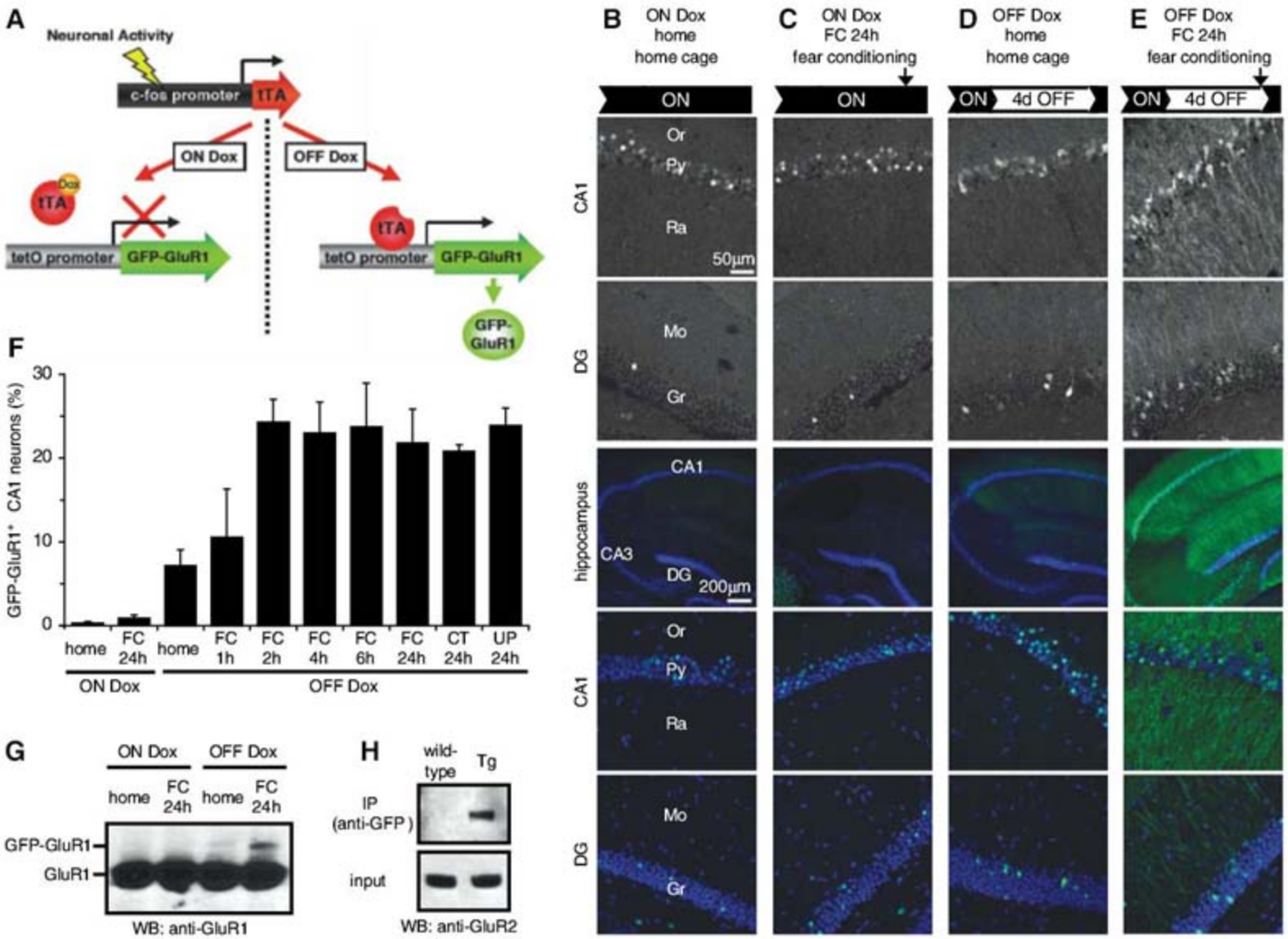


Fig. 1. Basic features of GFP-GluR1$^{c\text{-}fos}$ Tg mice. (**A**) Schematic representation of the transgenic system. The *c-fos* promoter was used to drive rapid and transient expression of the tetracycline-regulated transactivator (tTA) in activated neurons. The tTA in turn activates transcription of a tetO promoter-linked GFP-GluR1 in a Dox-regulated manner. (**B** to **E**) Confocal microscopy images of intrinsic GFP fluorescence (upper two panels) and GFP immunoreactivity (lower three color panels) in the hippocampal slices. Nuclei were stained with TO-PRO-3 (blue). Or, stratum oriens; Py, stratum pyramidale; Ra, stratum radiatum; Mo, molecular layer; Gr, granule cell layer; DG, dentate gyrus. (**F**) Proportion of CA1 neurons with cytoplasmic GFP immunoreactive signals. Data in this and all subsequent figures are represented as mean ± SEM. ON Dox: $n = 5$ mice for home and FC24h. OFF Dox: $n = 5$ for home; $n = 4$ for FC1h, FC4h, FC24h, CT24h, and UP24h; $n = 3$ for FC4h and FC6h. (**G**) Western blot analysis showing the expression of GFP-GluR1 and endogenous GluR1 in the hippocampus. (**H**) Coimmunoprecipitation of GFP-GluR1 and endogenous GluR2 in the hippocampus.

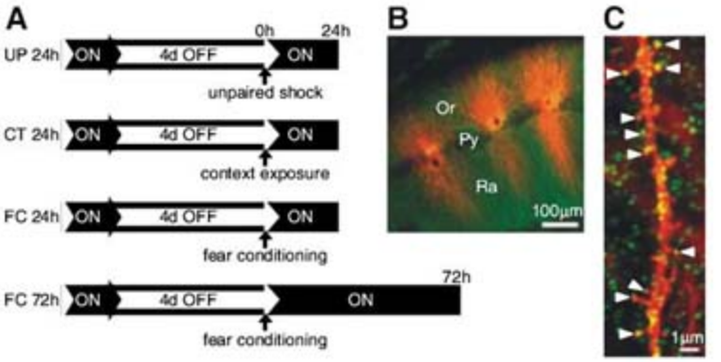
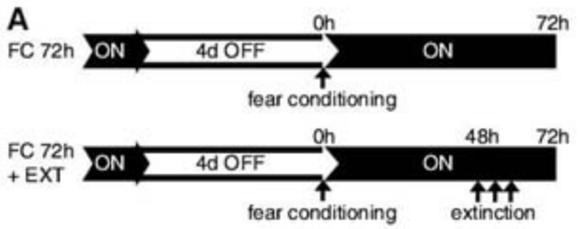


Fig. 2. Preferential recruitment of newly synthesized GFP-GluR1 to mushroom spines 24 hours after fear-conditioned learning. (**A**) Experimental design. (**B**) Confocal image of a hippocampal slice labeled with antibody to GFP (green) and DiI (red). (**C**) Confocal image showing surface GFP-GluR1 localization on apical dendrite after fear conditioning. Some spines were GFP positive (arrow heads) but some were negative, showing an uneven synaptic trafficking of newly synthesized GFP-GluR1. (**D**) Representative images of spine morphology and surface GFP-GluR1 localization. Scale bar, 1 μm. (**E**) Relative immunoreactivity from Western blot analysis (CT24h: $n = 4$, FC24h: $n = 3$). (**F**) The percentage of GFP-GluR1$^+$ mushroom spines was significantly higher (apical: $F_{2,16} = 8.86$, $P = 0.0026$; basal: $F_{2,16} = 15.43$, $P = 0.0002$) in FC24h group (apical: $n = 7$ mice, 1927 spines; basal: $n = 7$ mice, 1295 spines) as compared with the CT24h group (apical: $n = 6$ mice, 1626 spines; basal: $n = 6$ mice, 1292 spines) and the UP24h group (apical: $n = 6$ mice, 2240 spines; basal: $n = 6$ mice, 1974 spines). The proportion was significantly decreased (apical: $P = 0.0056$; basal: $P = 0.0010$; t-test) at 72 hours after conditioning (apical: $n = 6$ mice, 1233 spines; basal: $n = 6$ mice, 1138 spines). $^*P < 0.01$.



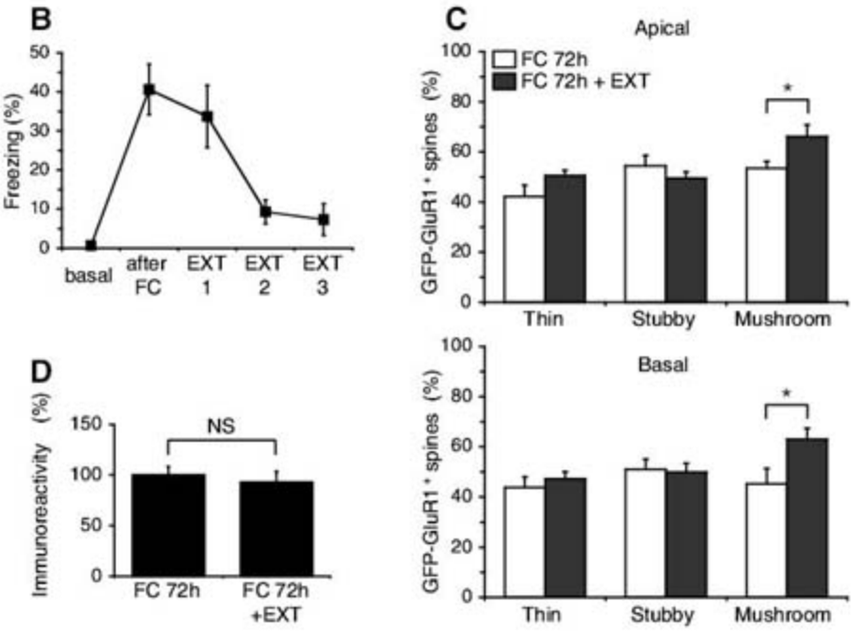


Fig. 3. Distribution of preexisting GFP-GluR1 after memory-extinction training. (**A**) Experimental design. (**B**) The percentage of time spent freezing before fear conditioning (basal), after fear conditioning (after FC), and during three trials of extinction training (EXT1~3); $n = 7$. (**C**) The proportion of GFP-GluR1$^+$ mushroom spines was significantly higher (apical: $P = 0.045$; basal: $P = 0.033$; t-test) in the extinction trained group (apical: $n = 7$ mice, 1870 spines; basal: $n = 6$ mice, 1845 spines) compared with the control group (FC72h, the same data as in Fig. 2F). (**D**) Relative immunoreactivity from Western blot analysis ($n = 3$ for each condition).

time points specifically in behaviorally activated neurons.

Dendritic spines are morphologically diverse but typically categorized into thin, stubby, and mushroom type (*21*, *22*) (Fig. 2D). We counted the proportion of surface GFP immunopositive spines (GFP-GluR1$^+$ spines) on apical and basal dendrites in the trained group (FC24h) and the two control groups (CT24h and UP24h) and analyzed the data separately for each spine type (Fig. 2F). There were no significant differences in GFP-GluR1$^+$ thin or stubby spines between groups. In contrast, for mushroom spines, we found a significant increase in the proportion of GFP-GluR1$^+$ spines in the fear-conditioned group compared with the context-only and unpaired-shock groups. The increased labeling returned to control levels by 72 hours after training (FC72h). Although GFP-GluR1 was delivered to ~50% of all spines, independent of behavioral contingencies or spine type, the presence of the shock reinforcer in the fear-conditioned group allows increased capture of receptor specifically in the mushroom-type spines.

Fear memory can be attenuated by extinction training, repeated exposure to an unreinforced conditioned stimulus. Extinction is thought to involve new learning rather than loss of the original memory, although the underlying mechanism remains elusive (*23*). We tested whether memory extinction alters the distribution of preexisting GFP-GluR1 in neurons that were activated during the original fear conditioning. GFP-GluR1$^{c\text{-}fos}$ Tg mice were withdrawn from Dox, fear conditioned, and returned to their homecage for 2 days on a high Dox diet. Conditioned mice received either three trials of extinction training (FC72h+EXT) (Fig. 3A) or were left undisturbed in the homecage (FC72h) (Fig. 3A). We confirmed a markedly reduced freezing response after extinction training (Fig. 3B). At 72 hours after fear conditioning, brain slices were prepared and the distribution of GFP-GluR1 was analyzed (Fig. 3C). For thin

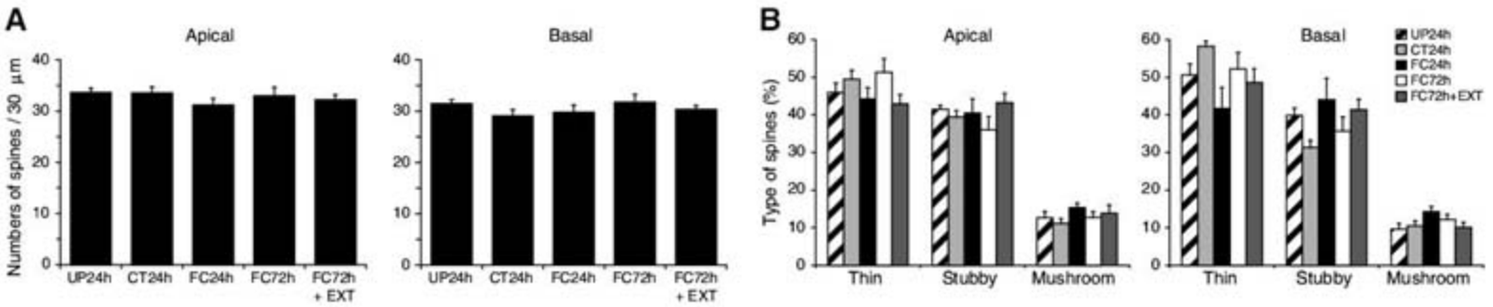


Fig. 4. Spine remodeling after behavioral training. (**A**) No significant changes were observed in spine density (apical: $F_{4,27} = 0.72$, $P = 0.58$; basal: $F_{4,27} = 0.88$, $P = 0.49$). $n = 6$ mice for UP24h, CT24h, and FC72h. $n = 7$ for FC24h and FC72h+EXT. (**B**) No significant changes were observed in spine-type distribution. $n = 6$ mice for UP24h, CT24h, and FC72h. $n = 7$ for FC24h and FC72h+EXT.

and stubby spines, extinction training did not affect the proportion of GFP-GluR1$^+$ spines, whereas the proportion of GFP-GluR1$^+$ mushroom spines was significantly higher in the extinction group than in the control group. A Western blot revealed that extinction training on Dox did not induce additional GFP-GluR1 expression (Fig. 3D).

We also tested whether fear conditioning induced changes in spine density, as is seen in some strong learning paradigms (*24*, *25*). We analyzed the proportion of each spine type and total spine density in behaviorally activated neurons. There were no significant changes in total spine density (Fig. 4A) or in spine-type distribution (Fig. 4B) in any group.

Our results demonstrate that fear conditioning alters trafficking of newly synthesized AMPARs in the adult hippocampus and that the change is spine-type specific. One of the most remarkable features of dendritic spines is their morphological diversity, with the mushroom spines considered to represent the most mature and stable spine morphology (*22*, *26*). At 24 hours after fear conditioning, these spines show an increased ability to incorporate and/or retain newly synthesized AMPARs. The change itself is not persistent and returns to baseline untrained levels by 72 hours. This may represent a subsequent change in receptor composition, for example, replacement of GluR1-containing receptors with GluR2/GluR3 receptors (*27*, *28*), or it may indicate a time-limited role for alterations of receptor composition in memory. Surprisingly, extinction training leads to a prolonged elevation in the proportion of GFP-GluR1$^+$ mushroom spines, even though it also leads to a reduction in the behavioral expression of the fear memory response. This could represent a redistribution of AMPARs to an alternate subset of synapses associated with extinction learning. Alternately, this could result from the stabilization of GFP-GluR1 in the synapses labeled with fear conditioning. Although little is known about the functional importance in the morphologically distinct spine classes in vivo (*22*, *26*), the current results demonstrate a unique role for the mushroom spines in learning-associated changes in receptor trafficking and suggest that they may be critical sites for behaviorally relevant plasticity. The learning-associated increase in GFP-GluR1 recruitment was detected in ~23% of mushroom spines, representing ~3% of all spines. Such changes are consistent with a sparse subset of synapses contributing to any given memory trace.

What role this receptor trafficking plays in the fear-conditioning paradigm is unclear. The trafficking changes require the same temporal pairing of context with shock required to produce the behavioral memory and likely represent an associative synaptic event. In fear conditioning, the hippocampus is generally thought to encode a representation of the context with the associative changes for fear occurring in the amygdala (*13*). However, electrophysiological studies indicate that the shock causes a partial remapping of the hippocampal representation (*29*), and the receptor trafficking changes may alter those synapses that participate in this incorporation of the unconditioned stimulus signal into the hippocampal representation.

LTP induction is known to increase AMPAR synthesis (*9*). By coupling expression of GFP-GluR1 to neuronal activity, we specifically examined this pool of newly synthesized receptor. In this paradigm, the earliest possible time point for the delivery of GFP-GluR1 to spines is offset from the behavioral training by several hours (see fig. S1). This suggests that at the time of learning there are changes in some spines that allow the capture of newly synthesized AMPARs at later time points. This type of "synaptic tagging" has been demonstrated to play a role in the maintenance of LTP (*8*). The current experiments demonstrate a similar mechanism operating during behavioral learning in the adult brain, which may contribute to the stabilization of long-term memory. They also implicate GluR1-containing AMPARs as one of the cargo molecules selectively delivered to tagged synapses.

References and Notes

1. R. Malinow, R. C. Malenka, *Annu. Rev. Neurosci.* **25**, 103 (2002).
2. T. Takahashi, K. Svoboda, R. Malinow, *Science* **299**, 1585 (2003).
3. S. Rumpel, J. LeDoux, A. Zador, R. Malinow, *Science* **308**, 83 (2005).
4. A. J. Heynen, E. M. Quinlan, D. C. Bae, M. F. Bear, *Neuron* **28**, 527 (2000).
5. D. R. Grosshans, D. A. Clayton, S. J. Coultrap, M. D. Browning, *Nat. Neurosci.* **5**, 27 (2002).
6. J. R. Whitlock, A. J. Heynen, M. G. Shuler, M. F. Bear, *Science* **313**, 1093 (2006).
7. Y. Dudai, *Curr. Opin. Neurobiol.* **12**, 211 (2002).
8. U. Frey, R. G. Morris, *Nature* **385**, 533 (1997).
9. A. Nayak, D. J. Zastrow, R. Lickteig, N. R. Zahniser, M. D. Browning, *Nature* **394**, 680 (1998).
10. M. Mayford *et al.*, *Science* **274**, 1678 (1996).
11. R. J. Smeyne *et al.*, *Neuron* **8**, 13 (1992).
12. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230 (2007).
13. S. G. Anagnostaras, G. D. Gale, M. S. Fanselow, *Hippocampus* **11**, 8 (2001).
14. S. H. Shi *et al.*, *Science* **284**, 1811 (1999).
15. Y. Hayashi *et al.*, *Science* **287**, 2262 (2000).
16. R. J. Wenthold, R. S. Petralia, J. Blahos II, A. S. Niedzielski, *J. Neurosci.* **16**, 1982 (1996).
17. A. L. Mammen, R. L. Huganir, R. J. O'Brien, *J. Neurosci.* **17**, 7351 (1997).
18. K. Archibald, M. J. Perry, E. Molnar, J. M. Henley, *Neuropharmacology* **37**, 1345 (1998).
19. H. Adesnik, R. A. Nicoll, P. M. England, *Neuron* **48**, 977 (2005).
20. S. Milanovic *et al.*, *Brain Res.* **784**, 37 (1998).
21. K. M. Harris, F. E. Jensen, B. Tsao, *J. Neurosci.* **12**, 2685 (1992).
22. E. A. Nimchinsky, B. L. Sabatini, K. Svoboda, *Annu. Rev. Physiol.* **64**, 313 (2002).
23. K. M. Myers, M. Davis, *Neuron* **36**, 567 (2002).
24. B. Leuner, J. Falduto, T. J. Shors, *J. Neurosci.* **23**, 659 (2003).
25. S. Knafo, G. Ariav, E. Barkai, F. Libersat, *Hippocampus* **14**, 819 (2004).
26. J. Bourne, K. M. Harris, *Curr. Opin. Neurobiol.* **17**, 381 (2007).
27. S. Shi, Y. Hayashi, J. A. Esteban, R. Malinow, *Cell* **105**, 331 (2001).
28. S. G. McCormack, R. L. Stornetta, J. J. Zhu, *Neuron* **50**, 75 (2006).
29. M. A. Moita, S. Rosis, Y. Zhou, J. E. LeDoux, H. T. Blair, *J. Neurosci.* **24**, 7015 (2004).

30. We thank S. Kupriyanov for pronuclear microinjection. This work was supported by NIH grants MH57368 and MH070020 (to M.M.). N.M. was supported by a postdoctoral fellowship from the Uehara Memorial Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1104/DC1
Materials and Methods
Figs. S1 to S4
References

31 August 2007; accepted 4 January 2008
10.1126/science.1149967

Rapid Neural Coding in the Retina with Relative Spike Latencies

Tim Gollisch* and Markus Meister†

Natural vision is a highly dynamic process. Frequent body, head, and eye movements constantly bring new images onto the retina for brief periods, challenging our understanding of the neural code for vision. We report that certain retinal ganglion cells encode the spatial structure of a briefly presented image in the relative timing of their first spikes. This code is found to be largely invariant to stimulus contrast and robust to noisy fluctuations in response latencies. Mechanistically, the observed response characteristics result from different kinetics in two retinal pathways ("ON" and "OFF") that converge onto ganglion cells. This mechanism allows the retina to rapidly and reliably transmit new spatial information with the very first spikes emitted by a neural population.

During natural vision, our gaze remains fixed for a mere fraction of a second. Sudden movements of the eye, called saccades, partition visual processing into short episodes (*1*, *2*). Each saccade exchanges the image that falls onto the retina; the new visual stimulus is then encoded into neural activity to be transmitted to the brain. Our visual system can analyze and classify a new complex scene in less than 150 ms (*3*), but the nature of the neural code that underlies this rapid visual processing has been elusive. Neurons in the vertebrate retina fire with remarkable temporal precision (*4*, *5*), so single spikes can, in principle, carry substantial information about visual stimuli. In order to assess how the retina transmits new visual information after a saccade, we investigated the responses of retinal ganglion cells to flashed visual images.

Spike trains were recorded simultaneously from many ganglion cells in the isolated salamander retina. The stimulus was a uniform gray field followed by appearance of a square-wave grating. Eight different shifted versions of the grating were used in a pseudo-random sequence. A ganglion cell typically responded to the appearance of the grating with a short burst of spikes (Fig. 1), and the vast majority of cells responded to most or even all of the stimuli. We characterized each burst by two numbers: the latency of the first spike after stimulus onset and the total spike count in the burst. For certain cell types, in particular fast and biphasic OFF cells (fig. S2), the spike count was very similar for all stimuli (Fig. 1B). By contrast, the spike latency for these cells varied across stimuli by as much as 40 ms (Fig. 1C). For repeats of the same stimulus, this latency was very reproducible, with a standard deviation of only 3 to 5 ms.

We calculated how much information the spike latency or the spike count conveys about which grating had been presented. Perfect identification of the stimulus among eight possibilities amounts to a maximum of 3 bits. The spike latency of a ganglion cell transmitted up to 2 bits of information on a single trial. The spike count provided considerably less information for the majority of all recorded cells (Fig. 1D). Subsequent brain regions may thus learn more about the stimulus from noting the time of the first spike after stimulus onset than by waiting for all spikes and noting the average firing rate.

In several sensory systems, shorter spike latencies result from stronger stimulation (*6*–*9*). This does not account for the present dependence of latency on spatial pattern. Stronger stimuli often generate higher spike counts, and indeed, gratings of higher contrast produced both more spikes and shorter latencies (fig. S3). By contrast, we observed a pronounced spatial tuning of the spike latency even when there were virtually no variations in spike count (Fig. 1); in some cases, shorter latencies even occurred in combination with fewer spikes (fig. S3).

Downstream brain centers can interpret the latency of a single neuron only if the onset time of the stimulus is known (*10*). If the new retinal image was initiated by an eye movement, then the brain does know the onset time, but it is unclear whether this motor information gets distributed to visual centers. We therefore asked what information can be extracted from visual signals alone by comparing latencies from neurons in the population (Fig. 2). For many pairs of ganglion cells, the difference between first spike times was strongly tuned with respect to the presented stimuli (Fig. 2C). In fact, the information contained in the latency difference reached values higher than 2 bits (Fig. 2D)—more than that from any single-cell absolute latency. One reason was the particular robustness of the latency difference to retinal noise. Each cell's latency underwent some trial-to-trial variation, but these fluctuations were often positively correlated in cells recorded simultaneously; when cell 1 fired earlier than usual, cell 2 tended to do the same (Fig. 2B). As a result, the latency difference (Fig. 2C) fluctuated less than expected from the noise in individual cell latencies (Fig. 2A). To assess the relevance of this compensation, we destroyed the noise correlations artificially by pairing the response of cell 1 with the response of cell 2 on the subsequent trial; this led to a substantial information loss of up to ~20% (Fig. 2D).

Stimuli of greater strength tend to produce shorter spike latencies in the sensory response. If the latencies of different neurons in the population are affected in similar fashion, downstream circuits might use the difference in spike latencies to extract stimulus quality independent of stimulus strength (*11*). We therefore presented the flashed gratings at different contrast levels. As expected, individual latencies increased at lower contrast. However, the shape of the latency tuning curve was well preserved at each contrast level (Fig. 2, E and F). Furthermore, the contrast-dependent shifts of the latency tuning curves were similar for different cells. As a result, the latency difference between two neurons was almost perfectly invariant to changes in contrast (Fig. 2G). In fact, a downstream decoder could recover almost all the spatial information without knowing anything about the contrast of the stimulus (Fig. 2H).

How can the observed latency code be explained in terms of neural mechanisms? We start by considering a standard framework for visual responses (*12*, *13*) and exploring its prediction for first-spike latencies. In this picture, the stimulus is first passed through a linear filter that summarizes retinal integration of the image over space and time (Fig. 3A). The "activation" signal emerging from the filter can be interpreted as the membrane potential of the ganglion cell. When this signal crosses a preset threshold, the model neuron fires a spike. For each ganglion cell, we measured the spatiotemporal filter in a separate reverse-correlation experiment (fig. S1) (*14*), whereas the threshold remained as a single free parameter. In using this model to process grating stimuli, one quickly finds that it cannot account for the observed responses. Because the stimulus is integrated linearly, a certain grating may elicit strong excitation, but then its sign-reversed counterpart will elicit inhibition and produce no spikes at all, counter to what was observed experimentally (Fig. 3, B and C).

Thus, one is forced to include nonlinear processing steps. Ganglion cells draw their excitatory input from bipolar cells (Fig. 3D). These have comparatively small receptive fields [<100 μm; (*15*)] and respond to light in essentially linear fashion (*16*), but transmission to retinal ganglion cells may involve a degree of rectification (*17*–*19*). Furthermore, bipolar cells come in two major types: ON bipolars are excited by an increase in light intensity and OFF bipolars by a decrease. Individual ganglion cells can receive inputs from both types (*8*, *20*–*24*). To include this structure of the inner retina into the model, we replaced the single spatiotemporal filter by a set of parallel filters that mimicked spatially local ON and OFF bipolar cells (Fig. 3E). Transmission from bipolar to ganglion cells was approximated by a half-wave rectifying function (Fig. 3F). Under these conditions, all stimuli led to

Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA.

*Present address: Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany.
†To whom correspondence should be addressed. E-mail: meister@fas.harvard.edu

excitatory activation of the ganglion cell (Fig. 3G). Moreover, the predicted timing of the first spike agreed remarkably well with the observed latency tuning curve (Fig. 3H) and outperformed alternative circuit schemes (fig. S6) for virtually all cells with substantial latency tuning (fig. S7).

Note that all the elements of this model are well-known components of retinal circuitry. Closer inspection reveals how the circuit accomplishes latency tuning. First, the rectifying synapses ensure that every stimulus excites the ganglion cell: any image change within the receptive field will activate some set of bipolar cells that transmit their excitation to the ganglion cell (*18*). The lasting and synchronous activation of both ON and OFF pathways by flashed gratings emphasizes their nonlinear summation and, thus, the need for separate filters in the model. By contrast, earlier studies of ganglion cells (*13*) were based on spatially homogeneous stimuli that only transiently activate one pathway at a time and, therefore, allow for simpler models (*14*). Second, the measured ON filters have slower kinetics than the OFF filters (Fig. 3E and fig. S5), such that ON stimuli affect ganglion cell spiking ~30 ms later than OFF stimuli. This is consistent with prior observations (*23*) and probably results from a signal transduction delay at the synapse between photoreceptors and ON bipolars (*25*, *26*). Thus, the proportion of light and dark stimulation within the receptive field determines the relative contribution of the ON and OFF pathways and modulates the time of the first threshold crossing.

This hypothesis for latency coding relies intimately on the convergence of parallel neuronal pathways with intrinsic kinetic differences. If this picture is correct, removal of one of the pathways should lead to a breakdown of latency tuning. We therefore exposed the retina to 2-amino-4-phosphono-butyrate, a metabotropic glutamate receptor agonist that blocks neural transmission to ON bipolar cells (*26*). The results were as predicted: Fast OFF ganglion cells ceased responding to about half of the stimuli (fig. S8), consistent with a loss of all the ON filters of the model as shown in Fig. 3F.

Although grating stimuli are convenient for systematic investigations, they do not capture the complex statistics of natural scenes. We thus briefly flashed a photographic image onto the retina (Fig. 4A). Across repeated presentations, the image was shifted to many different locations. In this way, spikes from a single ganglion cell could be used to simulate a population of identical neurons with different receptive field locations. Responses to the natural image resembled those to the gratings. For fast OFF cells, almost all image presentations elicited spike bursts that varied in latency by about 40 ms (Fig. 4B), and this latency was systematically related to the stimulus. Indeed, by simply plotting the differential spike latencies as a gray-scale code, we obtained a rather faithful neural representation of the raw visual image (Fig. 4C). This demonstrates the high quality of the latency information. Subsequent brain regions could use this for local image computations; for example, a neuron that detects spike coincidence among multiple ganglion cells would be selective for contour lines or edges in the image.

The corresponding neural image created from spike counts (Fig. 4D) is more blurred and noisy, and the highest values are observed near edges in the stimulus. In the flat regions, the center-surround antagonism of ganglion cell receptive fields (fig. S1) reduces firing activity. But because the effect of the receptive-field surround is delayed relative to the center (*20*), it does not affect the first spike in a burst. In fact, latency and spike count may serve to encode complementary stimulus features, which could support a rapid scene analysis with subsequent refinement (*27*). Furthermore, in natural vision, the ongoing fixational eye movements after a saccade may well affect the spike count throughout fixation, but should have negligible effects on the timing of the first spike.

Altogether, our results suggest that a population code based on differential spike latencies can be a powerful mechanism to rapidly transmit a new visual scene. Rapid saccades are ubiquitous in animal vision (*1*). In salamanders, they result from turns of the head and constitute a vital part of the approach to a prey (*2*). During a saccade, many ganglion cells are strongly suppressed (*19*, *28*), such that the first spike after a saccade is easily recognized. The differential latency of these

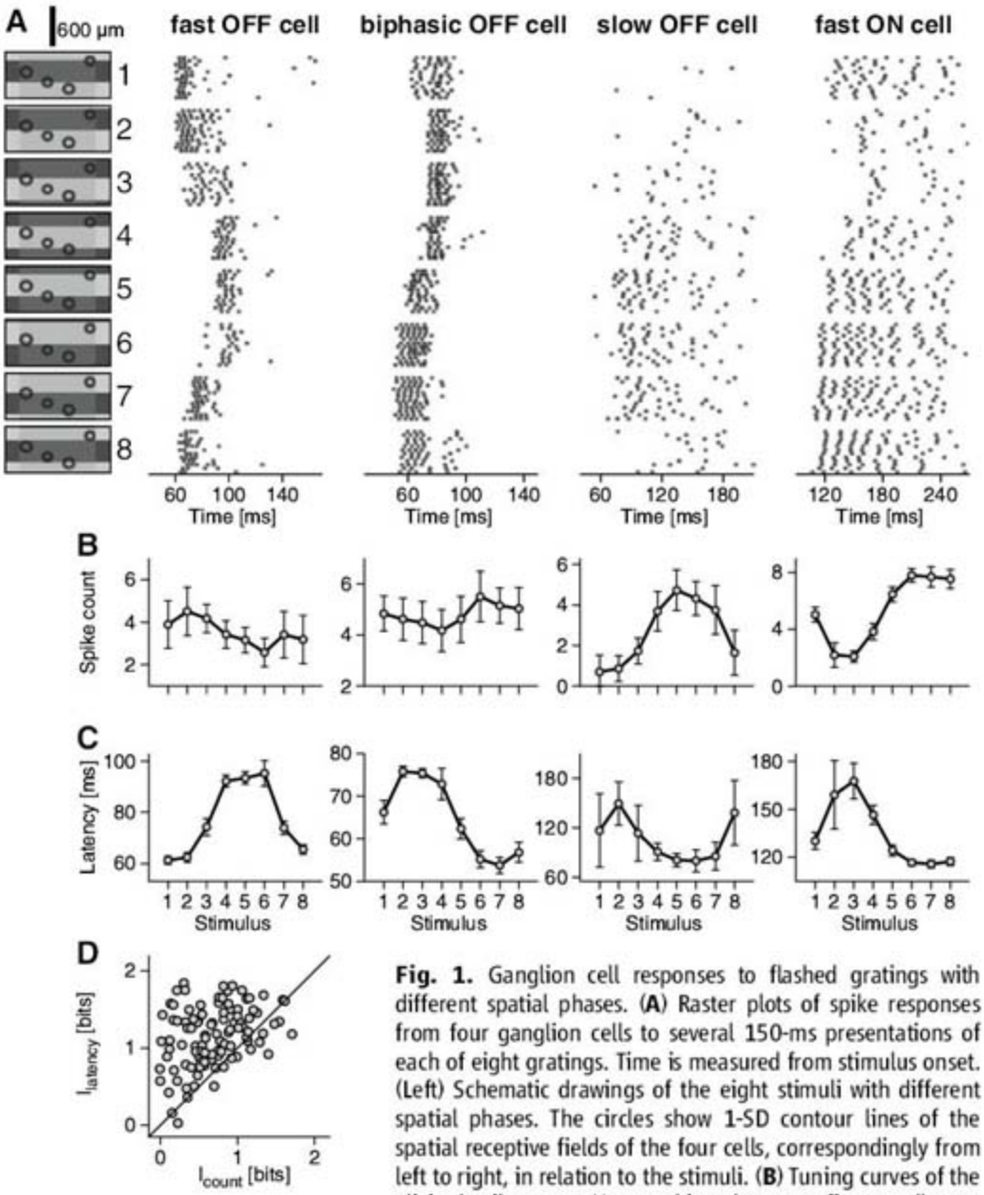


Fig. 1. Ganglion cell responses to flashed gratings with different spatial phases. (**A**) Raster plots of spike responses from four ganglion cells to several 150-ms presentations of each of eight gratings. Time is measured from stimulus onset. (Left) Schematic drawings of the eight stimuli with different spatial phases. The circles show 1-SD contour lines of the spatial receptive fields of the four cells, correspondingly from left to right, in relation to the stimuli. (**B**) Tuning curves of the elicited spike count. Here and in subsequent figures, all error bars show the standard deviation across trials with the same stimulus. (**C**) Tuning curves of the first-spike latency. "Fast OFF" and "biphasic OFF" cells typically showed strong tuning in the latency and only mild tuning in spike count; despite their names, these cell types receive input from both ON and OFF pathways (*19*). "Slow OFF" and "ON" cells, on the other hand, displayed good tuning in the spike count and often did not respond with spikes to all stimuli. The relatively long latencies are typical for cold-blooded animals. (**D**) Information about the stimulus identity contained in the spike count and in the latency, respectively, for all recorded cells. For a subdivision of the data by ganglion cell type, see fig. S2.

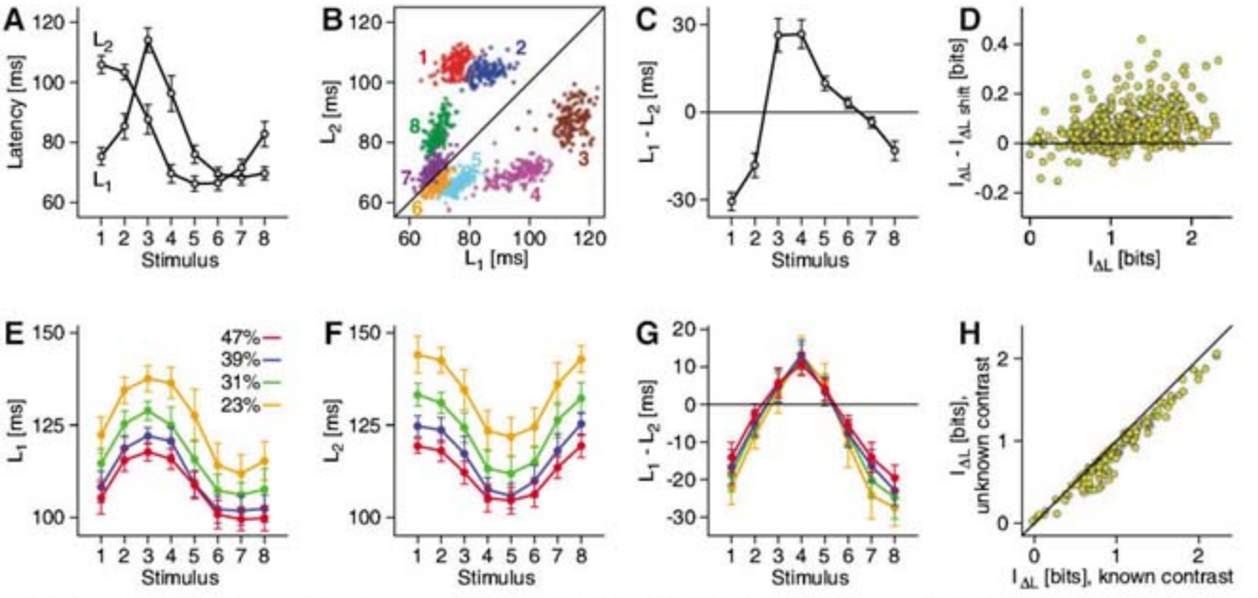


Fig. 2. (**A** to **D**) Encoding by relative latencies of pairs of ganglion cells. (A) Latency tuning curves for two simultaneously recorded fast OFF cells. (B) Scatter plot of latencies (L_1, L_2) for the two cells. The diagonally elongated distributions of the data show that L_1 and L_2 were positively correlated across trials with the same stimulus (*14*). (C) Tuning curve of the latency difference $L_1 - L_2$. (D) Information theoretical analysis of latency differences from simultaneously recorded cell pairs. The information $I_{\Delta L}$ about the stimulus contained in the latency difference $\Delta L = L_1 - L_2$ is plotted against the information loss $I_{\Delta L} - I_{\Delta L\ \mathrm{shift}}$ that occurred when L_2 was shifted by one trial with respect to L_1. (**E** to **H**) Contrast-invariant encoding by pairs of ganglion cells. (E) Latency tuning curves for a fast OFF cell whose responses were recorded for flashed gratings at different Michelson contrast levels. (F) Latency tuning curves of a second, simultaneously recorded fast OFF cell. (G) Tuning curves of the latency difference for the two neurons. (H) Information about the stimulus pattern, carried by latency differences whether the contrast level is known or not. All cell pairs that were recorded at the four different contrast levels were analyzed. The data points near the diagonal show that little information is lost by ignoring the stimulus contrast.

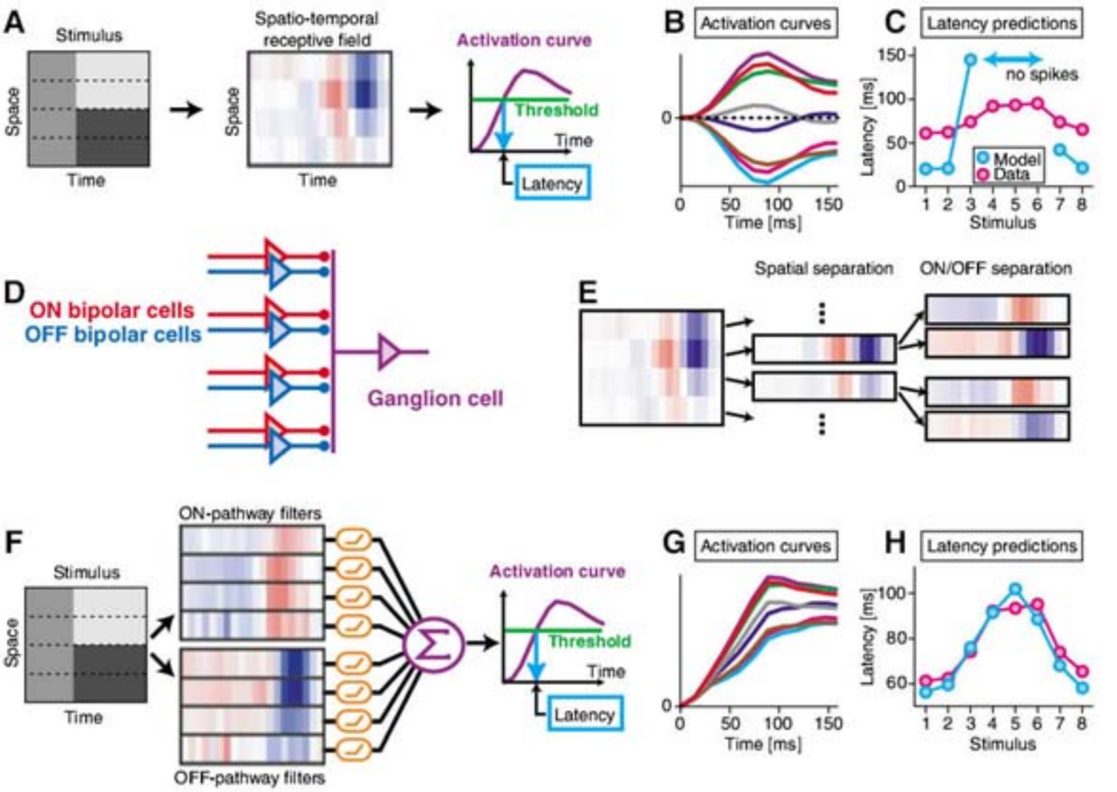


Fig. 3. Modeling the response latencies of retinal ganglion cells. (**A**) Standard framework for modeling ganglion cell responses. The stimulus (left) is gray illumination followed by a grating. This is convolved with a spatiotemporal filter (middle) representing the ganglion cell's receptive field (fig. S1). When the resulting activation curve exceeds a preset threshold, the first spike is fired (right). (**B**) Activation curves computed for each of the eight grating stimuli, by using the measured spatiotemporal filter for a sample fast OFF ganglion cell (first cell in Fig. 1). (**C**) Predicted and measured dependence of the latency on the stimulus. The threshold is the only free parameter of the model and was optimized from a χ^2 fit to the measured latency tuning curve. Several stimuli did not lead to positive activation and thus did not predict spikes. (**D**) Retinal interneuron pathways that motivate a revised model. Each small subregion of the receptive field activates both ON and OFF bipolar cells. The ganglion cell pools inputs across subregions and from both bipolar types. (**E**) Separation of the spatiotemporal filter into spatial subunits and subsequently into ON and OFF pathway contributions. [See (*14*) and fig. S5 for measurement of these contributions.] (**F**) Multi-pathway model of the response: Each subregion of the stimulus is passed through an ON filter and an OFF filter. The filter output is half-wave rectified and then pooled with all other outputs to yield the activation curve. (**G**) Activation curves computed for each of the eight grating stimuli, using the model in (F). Note that each stimulus produced excitation. (**H**) Measured latency tuning curves and predictions of the model in (F), after optimization of the threshold.

spikes encodes fine spatial detail (Figs. 1 and 4), yet it is almost entirely invariant to the overall stimulus contrast level (Fig. 2). Furthermore, it is robust to retinal noise (Fig. 2), and it provides information about the pattern in the shortest possible time, namely, with the very first spikes. Many vertebrates have specific ganglion cells, often multiple types, that combine inputs from both ON and OFF pathways (*20*–*23*), as evidenced by their ON/OFF response characteristics or by their dendritic morphology that connects to both ON and OFF bipolar cells. These neurons are candidate carriers of a latency code (Fig. 3). Synapses in the early visual pathway are very efficient, such that short spike bursts are reliably transmitted from retina to cortex (*29*). Moreover, certain neurons in visual cortex are exquisitely sensitive to the coincidence of spikes on their

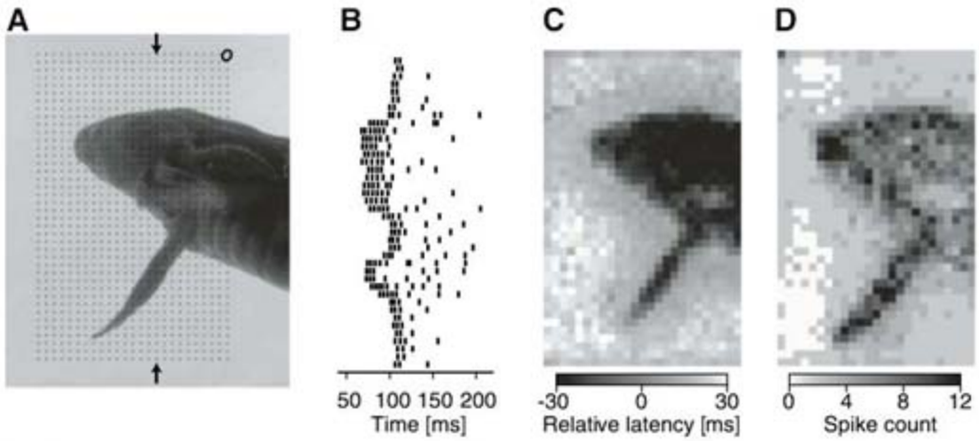


Fig. 4. Responses of a fast OFF ganglion cell to a flashed natural image. (For results from other cell types, see fig. S9.) (**A**) Photograph of a swimming salamander larva projected on the retina. The ellipse in the upper right corner shows a sample 1-SD outline of a ganglion cell receptive field. In each of 1000 presentations, the image was shifted slightly, and the grid of dots marks the resulting centers of the receptive field. Presentations were separated by gray illumination at the mean intensity of the photograph. The image onset produced luminance changes at most locations. (**B**) Spike trains of the ganglion cell for receptive-field locations along the column marked by the arrows in (A). (**C**) Gray-scale plot of the differential spike latency on single-trial presentations at the locations marked with dots in (A). The reference latency was chosen as the average value at all locations (*10*). (**D**) Corresponding gray-scale plot of the spike counts.

afferents (*30*), which is one possible readout mechanism for a latency code. Cortical neurons themselves carry substantial sensory information in their response latencies (*6*, *7*, *31*). Thus, it is conceivable that early aspects of sensory processing operate on the basis of the classification of spike latency patterns.

References and Notes

1. M. F. Land, *J. Comp. Physiol. A* **185**, 341 (1999).
2. C. Werner, W. Himstedt, *Zool. Jahrb. Abt. Allg. Zool. Physiol. Tiere* **89**, 359 (1985).
3. S. Thorpe, D. Fize, C. Marlot, *Nature* **381**, 520 (1996).
4. M. Meister, M. J. Berry, *Neuron* **22**, 435 (1999).
5. V. J. Uzzell, E. J. Chichilnisky, *J. Neurophysiol.* **92**, 780 (2004).
6. T. J. Gawne, T. W. Kjaer, B. J. Richmond, *J. Neurophysiol.* **76**, 1356 (1996).
7. D. S. Reich, F. Mechler, J. D. Victor, *J. Neurophysiol.* **85**, 1039 (2001).
8. M. Greschner, A. Thiel, J. Kretzberg, J. Ammermüller, *J. Neurophysiol.* **96**, 2845 (2006).
9. N. B. Sawtell, A. Williams, C. C. Bell, *Curr. Opin. Neurobiol.* **15**, 437 (2005).
10. S. M. Chase, E. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5175 (2007).
11. J. J. Hopfield, *Nature* **376**, 33 (1995).
12. E. J. Chichilnisky, *Network* **12**, 199 (2001).
13. J. Keat, P. Reinagel, R. C. Reid, M. Meister, *Neuron* **30**, 803 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. W. A. Hare, W. G. Owen, *J. Neurophysiol.* **76**, 2005 (1996).
16. S. A. Baccus, M. Meister, *Neuron* **36**, 909 (2002).
17. J. D. Victor, R. M. Shapley, *J. Gen. Physiol.* **74**, 671 (1979).
18. J. B. Demb, K. Zaghloul, L. Haarsma, P. Sterling, *J. Neurosci.* **21**, 7447 (2001).
19. M. N. Geffen, S. E. de Vries, M. Meister, *PLoS Biol.* **5**, e65 (2007).
20. F. S. Werblin, J. E. Dowling, *J. Neurophysiol.* **32**, 339 (1969).
21. F. M. de Monasterio, *J. Neurophysiol.* **41**, 1435 (1978).
22. F. R. Amthor, E. S. Takahashi, C. W. Oyster, *J. Comp. Neurol.* **280**, 97 (1989).
23. D. A. Burkhardt, P. K. Fahey, M. Sikora, *Vis. Neurosci.* **15**, 219 (1998).
24. J. L. Coombs, D. Van Der List, L. M. Chalupa, *J. Comp. Neurol.* **503**, 803 (2007).
25. J. F. Ashmore, D. R. Copenhagen, *Nature* **288**, 84 (1980).
26. X. L. Yang, *Prog. Neurobiol.* **73**, 127 (2004).
27. M. D. Menz, R. D. Freeman, *Nat. Neurosci.* **6**, 59 (2003).
28. B. Roska, F. Werblin, *Nat. Neurosci.* **6**, 600 (2003).
29. P. Kara, R. C. Reid, *J. Neurosci.* **23**, 8547 (2003).
30. W. M. Usrey, J. M. Alonso, R. C. Reid, *J. Neurosci.* **20**, 5461 (2000).
31. S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, M. E. Diamond, *Neuron* **29**, 769 (2001).

32. We thank F. Engert and members of the Meister laboratory for advice. This work was supported by grants from the National Eye Institute (M.M.) and the Human Frontier Science Program Organization (T.G.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5866/1108/DC1
Materials and Methods
Figs. S1 to S10
References

23 August 2007; accepted 11 December 2007
10.1126/science.1149639

Predicting Human Interactive Learning by Regret-Driven Neural Networks

Davide Marchiori[1] and Massimo Warglien[2]*

Much of human learning in a social context has an interactive nature: What an individual learns is affected by what other individuals are learning at the same time. Games represent a widely accepted paradigm for representing interactive decision-making. We explored the potential value of neural networks for modeling and predicting human interactive learning in repeated games. We found that even very simple learning networks, driven by regret-based feedback, accurately predict observed human behavior in different experiments on 21 games with unique equilibria in mixed strategies. Introducing regret in the feedback dramatically improved the performance of the neural network. We show that regret-based models provide better predictions of learning than established economic models.

The surge of interest in the neural bases of economic behavior (*1*–*3*) prompts the question of how well neural networks can model human interactive decision-making (*4*). This question implies two issues: the choice of the network architecture and the selection of input information to the network that has to be both economically and neurophysiologically motivated.

Interactive learning differs from individual learning in that, given n agents, each agent adapts to behaviors that are modified by the concurrent learning of the other $n-1$ agents. It has an obvious relevance in economic contexts, but (more generally) much of human learning that occurs in social contexts has an interactive nature. Experimental game theory has provided a large set of laboratory data on human interactive learning in repeated games (*5*), often contradicting the predictions of standard game theory. The need for models of interactive learning in games arises from the difficulties of ordinary game-solution concepts to explain both the trajectories and the long-run stationary state of experimentally observed human behavior in repeated games. Games with unique equilibria in mixed strategies are an especially interesting case, because Nash equilibrium not only fails to approximate behavior in early rounds but also is often a poor predictor of the stable behavior emerging in the long run.

Until now, two main modeling strategies have been used with some success in trying to fit and predict how humans learn in repeated games in a laboratory setting. One modeling strategy extends a classical paradigm of learning theory (i.e., rein-

[1]Interdepartmental Center for Research Training in Economics and Management (CIFREM), University of Trento, Italy. [2]Advanced School of Economics and Department of Business Economics and Management, Ca' Foscari University, Venezia, Italy.

*To whom correspondence should be addressed. E-mail: warglien@unive.it

forcement learning) (*6*–*8*) to games. The second strategy builds hybrid models that blend reinforcement learning with modeling the evolution of a player's beliefs about other players' moves: The relative weight of both learning processes depends on parameters that can be tuned, in turn, by experience (*9*, *10*). More recently, a model emphasizing post-decision regret for foregone payoffs as the driver of learning has also been proposed (*11*). Interest in neuroeconomics suggests that a different modeling strategy might be explored, with the use of neural networks as models of human interactive behavior. We chose to keep the model as simple as possible, using (despite its well-known computational limitations) one of the most elementary learning neural network architectures: the simple (one-layered) analog perceptron (*12*, *13*). At the same time, we modified the feedback process to take into account some elementary economic considerations (in accordance with both theoretical insights and empirical evidence). Our basic assumption was that learning is driven by a sort of "ex-post" rationalizing process (*14*): Individuals modify their behavior by looking backward to what might have been their best move, once they know what the other individual's move was. They adjust in the direction of such an ex-post best response. Furthermore, we hypothesized that the intensity of such directional change is proportional to a measure of regret: how much they have missed by not playing such move (*15*, *16*). This is consistent with recent neuroscience research on individual decision-making, showing that regret affects learning and that both neurophysiological and behavioral responses to the experience of regret are correlated to its amplitude (*17*, *18*).

Our model maps the structure of a strategic game onto a neural network in a very straightforward way, by having an input node x_j corresponding to each payoff in the game matrix and by also including the opponent's payoffs and an output node y_i for each action available to a player k (Fig. 1). The input information is coded by having each input node take the value of the corresponding payoff in the current game; the output node activation is computed by summing up inputs to each output node weighted by the value of the incoming connections w_{ij} and transforming the summation via the hyperbolic tangent (tanh) activation function

$$y_i = \tanh\left(\beta \times \sum_j w_{ij} x_j\right) \quad (1)$$

where β is the parameter tuning the steepness of the tanh function.

The activation values of the output nodes can be interpreted as propensities to play an action and are transformed into actual probabilities of play by normalization.

Thus far, this model is a very conventional, simple analog perceptron, where learning is modeled, as usual, as adaptive updating of the connections' weights. We adopted a variant of the Hopfield update rule (*12*, *13*), which provides a more direct probabilistic interpretation of the classical perceptron learning procedure

$$w_{ij}^{t} = w_{ij}^{t-1} + \Delta w_{ij} \quad (2)$$

given the action m chosen by player k, a_m^k

$$\Delta w_{ij} = \lambda^2 \times [t_i(a^{-k}) - y_i] \times R^k(a_m^k, a^{-k}) \times x_j \quad (3)$$

where $t_i(a^{-k})$ is the ex-post best response of player k to the other players actions a^{-k}; y_i is its propensity to play action i; $R^k(\cdot)$ is the regret given the action a_m^k and other players' actions a^{-k}; x_j is the strength of the input to the node and can be interpreted as payoff saliency; and λ is the learning rate. Regret is computed as the difference between the actual payoff received by a player k and the maximum payoff obtainable, given other players' actions. Thus, the psychological intuition underlying Eq. 3 is that connection weight adjustment is driven by a series of factors that can be summarized as adjustment = learning rate × distance from ex-post best response × regret × input saliency.

As compared with Hopfield's perceptron rule, the main difference of this variant is that the error feedback is multiplied by the regret size. One version of the model (henceforth PB1) squeezes the number of parameters to one by equating β to λ. This choice is justified because the effects of both parameters on the adjustment process are highly correlated (computer simulations have confirmed that this parameter trimming implies no substantial

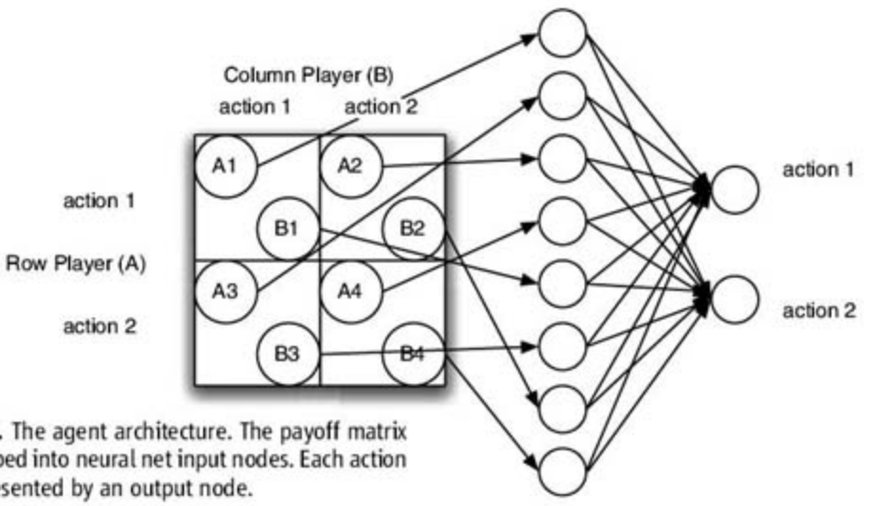


Fig. 1. The agent architecture. The payoff matrix is mapped into neural net input nodes. Each action is represented by an output node.

Average MSD scores (SD)	Average AICc scores (SD)								
2.45 (1.96)	–37.790 (18.13)	NNET2							
2.55 (1.78)	–41.380 (15.55)	Random	1, 15 17, 4						
1.65 (1.48)	–42.019 (15.93)	REL	13, 8 16, 6	15, 5 11, 10					
2.74 (2.85)	–42.313 (13.51)	Nash Eq.	8, 13 12, 9	10, 11 10, 11	8, 13 8, 13				
1.83 (1.09)	–43.464 (17.74)	BR	13, 8 16, 5	14, 6 11, 10	6, 14 13, 8	10, 9 11, 10			
1.07 (1.25)	–51.725 (18.33)	stEWA	15, 6 18, 3	16, 4 18, 2	17, 4 20, 1	14, 7 15, 6	18, 3 16, 5		
0.78 (0.66)	–52.550 (16.97)	PB1	16, 1 21, 0	21, 0 19, 2	17, 3 18, 2	18, 3 15, 6	20, 1 17, 4	14, 6 14, 7	
0.50 (0.35)	–52.681 (16.92)	NFP	20, 0 20, 1	21, 0 18, 3	19, 2 16, 5	18, 3 15, 6	20, 1 18, 3	18, 3 12, 9	15, 5 11, 10
0.86 (0.75)	–54.660 (17.14)	PB0	18, 2 21, 0	20, 1 20, 1	16, 5 19, 2	17, 4 15, 6	19, 1 20, 1	12, 9 14, 7	8, 12 15, 6	4, 17 12, 9
			NNET2	Random	REL	Nash Eq.	BR	stEWA	PB1	NFP

Fig. 2. Results. The first and second columns indicate average MSD and average AICc scores, respectively, over all 21 games. The third column and bottom row indicate model names. Cells in the remaining columns contain two sets of paired values: The top pair indicates the number of tasks for which the MSD score of the "row" model was significantly better or worse (the first and second numbers in each pair, respectively) than that of the "column" model, and the bottom pair shows those relationships for AICc scores. As a result of ties, the sums of each pair in a cell may be less than 21.

losses in descriptive and predictive performances). We also introduced a second model (PB0), which has no free parameters as a result of a simple form of meta-learning (*10*, *19*), allowing the endogenous determination of the learning rate λ. Once more, regret plays a central role: We assumed that the learning rate is tuned by a cumulative regret function that increases λ as the current ratio of experienced regret to the maximum possible regret exceeds the average and decreases λ in the opposite case

$$\lambda_t^k = \frac{\sum_t R_t^k}{\sum_t \max(R_t^k)} \quad (4)$$

where t is the number of iterations, R_t^k is the regret that is actually experienced by player k at round t, and $\max(R_t^k)$ is the maximum possible regret that player k could experience at time t (of course, in a repeated game, such value is constant).

To test the descriptive and predictive accuracy of our model, we considered experiments on 21 different games with unique equilibria in mixed strategies (*8*, *20*–*25*). Over more than 50 years, these experiments have been conducted by researchers other than the authors of this paper. The games have several actions, which range from two to five, that are available to each player. Of those games, 17 are constant sum games, whereas the remaining 4 are games in which players have no incentive to favor the other player: In other words, in each experiment, players had to learn strategies of conflict. In order to let learning processes unfold, we selected experiments with the constraint that there should be a minimum of 100 iterations of the stage game.

Our focus on such a class of games was motivated by multiple considerations. Because such games have unique equilibria, game theory lends a unique prediction of agents' behavior, providing a nonequivocal benchmark. Furthermore, all the 21 games that we considered have nondegenerate solutions: In equilibrium, subjects have to randomize their behavior. This is a source of cognitive complexity, despite the apparent simplicity of the game structures. Nash equilibrium turns out to be a poor predictor of observed behavior in such games (in many of them, it performs even worse than a "random behavior" prediction). Finally, this is the class of games on which the largest set of experiments with sufficient iterations is available.

We compared our model with different breeds of models (*7*): in particular, we took the Nash equilibrium, blind random behavior, and three of the most established learning models in the behavioral game-theory literature [i.e., the Basic Reinforcement Learning (BR) model (*7*), the Reinforcement Learning (REL) model by Erev *et al.* (*8*), and the self-tuning Experience-Weighted Attraction (stEWA) model (*10*)] as competitors, as well as the recent Normalized Fictitious Play (NFP) model (*8*, *11*). To single out the value added by introducing a regret term in the perceptron feedback, we further compared our model with the corresponding one-layer analog perceptron (NNET2) that uses the ordinary error feedback measure (dropping the regret term from Eq. 3) and has independent λ and β free parameters.

Given the availability of 21 different experimental conditions, it is appropriate to use—for each single condition (game)—the other 20 games to calibrate free parameters and predict behavior in the given condition (*6*, *7*, *26*). This provides 21 different predictions: one for each game. We used Mean Square Deviation (MSD) as a measure of goodness-of-fit for calibration and prediction (*27*). Although PB0 has no free parameters to estimate, PB1, reinforcement learning, and stEWA have one free parameter to estimate, and the REL model and NFP have two free parameters: Thus, the models to be compared differ both in functional form and in the number of free parameters. Among generally accepted criteria for comparing models with different complexity, we used the corrected Akaike's Information Criterion (AICc) (adjusted for sample size), which is the method according the lowest penalty to excess parameters.

Figure 2 shows the results of our analysis. The regret-based perceptrons PB1 and PB0 had the second- and third-best average MSD scores on all 21 prediction tasks, and in most conditions, were predicting better than other models, with the only exception of the other regret-based model, NFP (which was the best performer in terms of MSD). Notably, the no-parameter PB0 preserved much of the one-parameter PB1 performance, with average MSD scores much smaller not only than other nonparametric models but also than the parametric ones, with the obvious exception of PB1 and NFP. Once the number of free parameters was taken into account, however, the no-parameter perceptron PB0 had the lowest AICc score and compared favorably to all other models in most of the games. Thus, no matter how one measures performance, regret-based models always fared better than the other models, although PB0 gained an advantage from its great parsimony (*28*).

Furthermore, the PB0 and PB1 models clearly outperform the traditional NNET2 analog perceptron, demonstrating the determinant role played by the introduction of regret as a source of feedback for learning.

Another important advantage for models such as PB1 and PB0 comes from the nature of the learning tasks that can be modeled. Most human interactive learning happens in contexts where tasks do not repeat themselves identically over time, as in the experiments considered here. Generalization from examples and the learning of conditional behavior (different responses to different inputs) are natural features of human behavior. Standard models of economic learning (including the recent NFP) do not capture such features, because there is no way that they can model dependence of behavior from the perception of different game structures. On the contrary, even simple neural networks, such as those investigated here, can easily model generalization and conditional behavior and thus are open to the investigation of more realistic interactive learning tasks.

References and Notes

1. C. F. Camerer, *Science* **300**, 1673 (2003).
2. M. Kosfeld *et al.*, *Nature* **435**, 673 (2005).
3. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, *Nature* **441**, 876 (2006).
4. D. Sgroi, D. J. Zizzo, *Physica A* **375**, 717 (2007).
5. C. F. Camerer, *Behavioral Game Theory* (Princeton Univ. Press, Princeton, NJ, 2003).
6. A. Roth, I. Erev, *Games Econ. Behav.* **8**, 164 (1995).
7. I. Erev, A. Roth, *Am. Econ. Rev.* **88**, 848 (1998).
8. I. Erev, A. Roth, L. Slonim, G. Barron, *Econ. Theory* **33**, 29 (2007).
9. C. F. Camerer, T. H. Ho, *Econometrica* **67**, 827 (1999).
10. T. H. Ho, C. F. Camerer, J. K. Chong, *J. Econ. Theory* **133**, 177 (2007).
11. E. Ert, I. Erev, *J. Behav. Decision Making* **20**, 305 (2007).
12. J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).
13. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8429 (1987).
14. R. Selten, R. Stöcker, *J. Econ. Behav. Organ.* **7**, 47 (1986).
15. R. Selten, K. Abbink, R. Cox, *Exp. Econ.* **8**, 5 (2005).
16. S. Hart, A. Mas-Collel, *Econometrica* **68**, 1127 (2000).
17. N. Camille *et al.*, *Science* **304**, 1167 (2004).
18. G. Coricelli *et al.*, *Nat. Neurosci.* **8**, 1255 (2005).
19. D. Lee, *Nature* **441**, 822 (2006).
20. P. Suppes, R. C. Atkinson, *Markov Learning Models for Multiperson Interactions* (Stanford Univ. Press, Stanford, CA, 1960).
21. D. Malcolm, B. Lieberman, *Psychon. Sci.* **12**, 373 (1965).
22. B. O'Neill, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2106 (1987).
23. A. Rapoport, R. B. Boebel, *Games Econ. Behav.* **4**, 261 (1992).
24. J. Ochs, *Games Econ. Behav.* **10**, 202 (1995).
25. J. Avrahami, W. Guth, Y. Kareev, *Theory Decision* **59**, 255 (2005).
26. J. R. Busemeyer, Y. M. Wang, *J. Math. Psychol.* **44**, 171 (2000).
27. R. Selten, *Exp. Econ.* **1**, 43 (1998).
28. In a recent comparison of different learning models (*7*), NFP was tested on 10 of the games included in our data set, estimating its parameters on a different set of data. In this case, NFP did not perform better than the other models, including REL and reinforcement learning, confirming that (with two free parameters) the choice of the data set on which parameters are estimated may be a source of instability of the predictive performance of the model.

29. I. Erev, J. Ochs, and B. O'Neill kindly provided to us their original data sets. Discussions with R. Selten were a major source of inspiration for the project. P. Pellizzari, C. Gilbert, the anonymous reviewers and the editor helped us to substantially improve the paper. Ministero dell'Università e della Ricerca projects Fondo per gli Investimenti della Ricerca di Base RBNE03A9A7 and Progetti di Ricerca di Interesse Nazionale 2005139342 provided financial support for the research.

Supporting Online Material

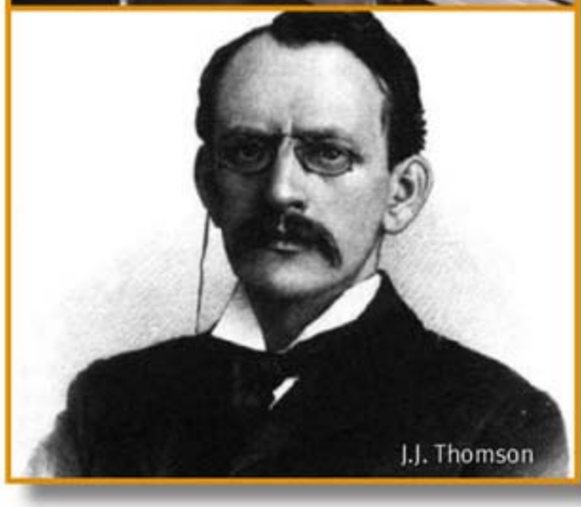
www.sciencemag.org/cgi/content/full/319/5866/1111/DC1
Materials and Methods
Figs. S1.1 to S1.10 and S2.1 to S2.37
Tables S1 to S7
References

1 October 2007; accepted 11 January 2008
10.1126/science.1151185

MASS SPECTROMETRY FOR THE MASSES

These days everyone seems to be doing it—mass spectrometry, that is. As protein chemists, molecular biologists, and even crime scene investigators rush to adopt this once-esoteric technology, the basic science revolution that made mass spectrometry user-friendly is striving to make it even more powerful.

By Alan Dove


J.J. Thomson

The revolution in mass spectometry (MS) started in places like Carlos Lebrilla's laboratory at the **University of California at Davis,** California. About 15 years ago, when the young chemist was a new faculty member at the school, he hung a sign on his most expensive piece of equipment. The sign, still hanging today, boldly advertised "Mass Spectrometry for the Masses." It was not merely the expression of an inside joke—it was more like a mission statement.

"When I started my career, I saw mass spectrometry going the same way that NMR went," says Lebrilla, who is now a full professor. Initially used only by physical chemists to determine molecular structures, nuclear magnetic resonance (NMR), rebranded as MRI, is now nearly as ubiquitous in clinical settings as the X-ray. If it follows the same path, mass spectrometry, an astonishingly fast, sensitive method for identifying individual molecules in complex mixtures, could revolutionize everything from clinical diagnostic testing to crime scene investigation. Unfortunately, the technique has long been considered too expensive, too finicky, and too difficult for deployment outside of specialized spectroscopy laboratories.

Over the past few years, though, mass spectroscopists and equipment manufacturers have been chiseling away at those problems, and a new crop of relatively inexpensive, compact, user-friendly devices is finally bringing the method within reach of ordinary researchers. But a mass spectrometer is still a substantial purchase, so scientists considering bringing such a system into their labs should investigate all of their options first.

Getting Your Charge On

Since J.J. Thomson's invention of the technique at the turn of the 20th century, scientists have developed numerous variants of mass spectrometry, but all of them follow the same essential process. The device draws a sample into a component called the ionization source, which converts molecules in the sample into ions. The ions then enter a mass analyzer, which applies different electrostatic and magnetic forces to the ions in a vacuum, separating them according to their mass-to-charge (m/z) ratios. Finally, the mass analyzer reports these ratios to a connected computer, which graphs and analyzes the data. From the results, a spectroscopist can identify specific molecular structures. That is, assuming nothing malfunctions.

"The one thing that anybody who has performed mass spectrometry experiments is well aware of is mass spectrometers tended to break fairly often," says Gary Siuzdak, director of the **Scripps Center for Mass Spectrometry** in La Jolla, California. However, he adds that with newer devices, failures in the core components of the mass spectrometer are relatively rare. Instead, faults seem to now occur more often in the liquid chromatography (LC) system that is usually coupled to the spectrometer for sample preparation and sometimes fails due to the complexity and nature of the samples being injected. "You have to be aware of what the limitations are of the instrument, what you can put into it," says Siuzdak, who also wrote continued >

> "You have to be aware of what the limitations are of the instrument, what you can put into it."

Look for these **Upcoming** Articles

Proteomics 1 — March 7
Biomarker Discovery — March 28
Genomics 1 — April 4

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, *nor is it meant to imply that their products or services are superior to those of other companies.*

CREDIT: (TOP) COURTESY OF PACIFIC NORTHWEST NATIONAL LABORATORY

"It's not simply the technology or the instrument that's going to be important; it's going to be the support structure in getting the instrument installed, training the people on the instrument and the ongoing challenges."

the standard introductory textbook for biological mass spectrometry (*The Expanding Role of Mass Spectrometry in Biotechnology*, 2006).

Even with careful technique, though, the machine could crash, highlighting one of the key features for which newcomers to mass spectrometry should shop. "It's not simply the technology or the instrument that's going to be important; it's going to be the support structure in getting the instrument installed, training the people on the instrument and the ongoing challenges," says Dave Hicks, senior director of the proteomics business group at **Applied Biosystems** in Foster City, California. Like other mass spectrometer makers, Applied Biosystems provides multiple levels of support with all of its devices, from field technicians who can fix common glitches to staff scientists who can answer high-level questions about experimental design and data analysis.

Besides a deep support network, researchers should look for the specific features they need, based on the types of samples they'll be analyzing. Though there are several kinds of ion sources and mass analyzers, workers in different fields tend to coalesce around specific arrangements.

In basic proteomics research, matrix assisted laser desorption/ionization (MALDI) has become the predominant ion source, and the mass analyzer is often an ion trap. MALDI is good for ionizing highly complex samples from protein-friendly buffers, and the ion trap can eject specific ion species from the sample seriatim, until all of the ions have been analyzed. It's a particularly good arrangement for researchers who aren't sure what they're looking for, and need to capture everything.

Scientists working in a more applied setting, such as a clinical lab, usually need to process samples more quickly than their basic research colleagues, but may only need to detect a comparatively small, defined set of analytes. For these users, electrospray ionization may be a good choice, combined with a triple-quadrupole mass analyzer. The triple-quadrupole design isn't quite as versatile as the ion trap, but it is efficient and inexpensive.

Less Than a Lamborghini

Of course, the term "inexpensive" is relative. For an entry-level triple-quadrupole mass spectrometer, complete with a liquid chromatography system capable of handling a particular class of samples, expect to spend around $100,000. Adding more features, more speed, or more versatility can quickly double that figure.

A six-digit purchase will certainly give most researchers pause, especially in grant-starved academic labs, but the prices on the latest entry-level mass spectrometers are still a historical bargain. "No matter what instrument you buy, you're most likely not going to lose your job because you bought that instrument," says Siuzdak. He adds that a decade ago, when even a cheap mass spectrometer could cost $300,000 and the technology was less reliable, a poor choice could indeed shorten one's career.

Nonetheless, scientists will still want to shop around before settling on a particular machine; there are many similar products to choose from, and each has its own strengths and weaknesses. "I would say that we're all competing head to head, but then the practical result is that we tend to have our respective areas of strength," says Ken Miller, director of marketing for the LC/MS group at **Agilent Technologies** in Santa Clara, California.

Agilent is a relatively recent entrant in the mass spectrometry field, but the company has quickly developed a large base of users. "I think we've been seeing dramatic growth across whole markets," says Miller, adding that the customers seem to be a mix of those migrating from other brands and those entering mass spectrometry for the first time. In response, Agilent has introduced several entry-level models for different types of users, as well as a series of high-end systems for experienced spectroscopists.

The same demands have driven longtime mass spectrometer maker **Thermo Fisher Scientific** in Waltham, Massachusetts, to expand its product lines. "Historically one of our key focuses has been on the high-end systems and the high-end users. But we're seeing this transfer of utilization of mass spectrometry to the non-mass-spec experts," says Lester Taylor, director of product marketing for mass spectrometry at the company. Accordingly, Thermo Fisher has introduced a series of new devices designed for these users. "There was a strong need to ensure that the operation and use of these systems is simplified, and that required [revising] the interface and the software that's used to run the system," says Taylor. Indeed, user-friendly software has become a major selling point for most

Mac, PC, or MS?

Besides an ionization source and a mass analyzer, every modern mass spectrometer also includes a computer, so it's no surprise that device makers spend a lot of time developing software. In theory, advanced mass spectroscopists could use other programs to collect and interpret the raw data, but in practice few do.

"Generally I think people use the customized package that comes with the system from the vendor, because obviously it's the operating system that drives the instrument, generates the data you need, and is best suited in many applications for processing it," says Lester Taylor, director of product marketing for mass spectrometry at Thermo Fisher Scientific of Waltham, Massachusetts. However, Taylor adds that many users also export the processed data to other applications for further analysis.

Both experienced and entry-level mass spectroscopists may want to integrate their mass spectrometry data with other quantitative experiments in their labs. Those who already use a laboratory information management package, such as Thermo Fisher's popular WatsonLIMS, will find that most vendors' mass spectrometers will export data to the system fairly easily. In proteomics, researchers often prefer to send the data to Mascot, a database search tool from **Matrix Science.** Mascot uses the spectrometer's data stream and online primary protein sequence databases to identify proteins in a sample quickly. Researchers can also plug the Mascot data into yet another application, **Proteome Software's** Scaffold, to get even more refined identification and analysis. Several other software companies also offer complementary products.

CREDIT: MACIEJ KOTLINSKI

entry-level mass spectrometers (see "Mac, PC, or MS?").

While the streamlined interfaces certainly simplify training—Siuzdak recently taught a high school student to do mass spectrometry experiments during a rotation in his lab—advances in the fundamental technology of mass spectrometry have made the initial purchase more complex. "Even though many people can use them, you need quite a bit of expertise in helping you pick out the one that would best suit your needs," says UC Davis's Lebrilla. He adds that while salespeople are a good source for technical information, prospective buyers should also talk to other researchers in their field who have bought mass spectrometers. The American Society for Mass Spectrometry (www.asms.org) also provides extensive resources for newcomers.

Pocket-sized Power

The growing population of entry-level mass spectroscopists is a diverse crowd, as manufacturers have already discovered. "I represent what we call the proteomics mass spec side of the business," says Applied Biosystems' Hicks. In academic and pharmaceutical proteomics, many researchers are thinking of buying their first mass spectrometer, but it won't be their first exposure to mass spectrometry. "Many of these sites already have some sort of a core lab or a core mass spectrometry facility," Hicks explains, so the question is not whether to use the technology, but whether to bring it into one's own laboratory. Users in this category often need a fairly advanced system, and are probably less intimidated by complex hardware and software.

When advocates like Siuzdak and Lebrilla talk about bringing the technology to the "masses," though, they are usually referring to users who have a specific problem in mind, but little or no background in mass spectrometry. "We are definitely seeing good growth in these markets and increasing adoption of mass spectrometry," says Joe Anacletto, senior director of applied markets and clinical research at Applied Biosystems. Unlike Hicks's customers, Anacletto's often have mass spectrometry thrust upon them by regulatory changes.

Environmental testing, food science, and police forensic laboratories exemplify this trend. As the public, legislatures, and courts have learned more about environmental pollutants and trace evidence, they have raised their standards for detecting and analyzing specific compounds. "This drives these testing labs to look for newer technologies that can handle the broader array of things they're going to be looking for," says Anacletto.

Indeed, the advance of mass spectrometry technology in basic research labs will likely increase the number of important biological and environmental markers even further. With the current crop of robust, sensitive mass spectrometers, biomarker researchers have embarked on ambitious efforts to find new indicators of disease, often by comparing blood or serum samples from patients with and without a particular condition.

Separation Anxiety

Though the mass spectrometry has gotten easier in these studies, the sample preparation problem has been harder to address. Serum, for example, consists mainly of a handful of highly abundant proteins, which tend to swamp sensitive mass spectrometry assays and conceal the much scarcer proteins that might change during pathogenesis.

In response, several companies have introduced immunoaffinity columns specifically designed to remove these abundant proteins. The columns, first introduced around 2003 and now available from companies such as **Sigma-Aldrich** in St. Louis, Missouri, and Agilent, have also been refined for greater efficiency and selectivity.

Prefiltering serum samples can certainly simplify a biomarker study, but the columns also have some disadvantages. Because they add a step to the sample preparation, they inevitably deplete some molecules by nonspecific binding. If an important biomarker is extremely scarce, it might disappear entirely after serum filtration. Also, scientists hoping to develop assays in animal systems before moving to human trials may find their protocols hard to translate; relatively few immunodepletion columns are available for animal systems, as companies have focused on developing the more popular human serum columns.

The Matrix and Beyond

Serum is not the only place biomarker investigators have been searching, of course. For those more interested in characterizing the markers in solid tissues, a major challenge has been to determine which cells in a complex sample are actually producing particular proteins. One elegant option, recently introduced by mass spectrometry giant **Bruker Daltonics** of Billerica, Massachusetts, is an integrated tissue imaging, sample preparation, and spectrometry platform called ImagePrep. After fixing a layer of tissue to the system's stage, researchers can overlay a precisely calibrated matrix, dry the sample, and perform MALDI spectrometry in situ. The result is a complete picture of the tissue structure, overlaid with spectra of the proteins expressed in each area.

As the new gear helps generate waves of additional data on everything from biomarkers to environmental chemicals, experts expect demand to increase for even simpler mass spectrometry platforms, especially for use in the field or at a patient's bedside. "The next stage of that development is to make more portable mass spectrometers. There are a lot of other groups looking at miniaturization of mass spectrometers, so you can actually just carry it around with you," says Lebrilla. Spectrometry for the masses, indeed.

Alan Dove is a science writer and editor based in Massachusetts.

DOI: 10.1126/science.opms.p0800022

Featured Participants

Applied Biosystems
www.appliedbiosystems.com

Agilent Technologies
www.agilent.com

Bruker Daltonics
www.bruker.com

Matrix Science
www.matrixscience.com

Proteome Software
www.proteomesoftware.com

Scripps Center for Mass Spectrometry
masspec.scripps.edu

Sigma-Aldrich
www.sigma-aldrich.com

Thermo Fisher Scientific
www.thermofisher.com

University of California, Davis
www.ucdavis.edu

New Products

Centrifugal Sample Prep Kits

The new C18, C8, and C4 SpinTip Sample Prep kits contain tips and reagents for reversed-phase solid-phase extraction of peptide and protein samples for electrospray or matrix-assisted laser desorption/ ionization mass spectrometry. They are particularly useful when runs of larger amounts of solution are involved; they can process sample solutions of up to 1 mL and 200 μg peptide/ protein. Research applications of the kits include effective desalting and detergent removal from biomolecule samples and cleanup of peptides from in-gel digested proteins or of proteins electroeluted from polyacrylamide gels.

Protea Biosciences
For information 877-776-8321
www.proteabio.com

Two-Dimensional Electrophoresis System

A two-dimensional electrophoresis system offers the highest focusing voltage available for the first dimension and has an easy-to-use, highly reproducible second dimension separation tank. The complete system includes the IEF100 First-dimension Isoelectric Focusing (IEF) Unit and the SE900 Second-dimension Gel Electrophoresis Unit. The IEF100 helps researchers increase the quality and speed of their IEF separations. The SE900's single-tank design ensures leak-free runs and the ability to reuse buffer. The IEF100's integrated 12,000-volt power supply enables fast run times and provides both ethernet and RS232 ports for reporting and control. The first dimension IEF can be run with IPG (immobilized pH gradient) strips of any length up to 24 cm. Up to six IPG strips can be focused at one time and monitored individually. The IEF100 generates first-dimension separations, then the focused strips are transferred to an SE900 second-dimension slab gel for the second dimension separation by molecular weight. The separation tank can hold up to six second-dimension gels.

Hoefer
For information 800-227-4750
www.HoeferInc.com

Handheld FTIR Spectrometer

A new handheld, portable Fourier transform infrared (FTIR) spectrometer weighs only six pounds. This innovative analyzer can be operated using only one hand and is capable of handling attenuated transmission reflectance (ATR) applications for users in the general laboratory. Designed for sampling almost any material, the system features interchangeable internal reflectance and external reflectance sampling interfaces. Designed to analyze solids, pastes, gels, and liquids with a traditional ATR, it is also capable of analyzing large surfaces with a unique internal reflection sampling system. It combines all the capabilities of a laboratory-based FTIR with the advantages of a portable instrument.

A2 Technologies
For information 203-312-1106
www.a2technologies.com

Metabolite Identification Software

MetWorks 1.1.0 is an updated version of metabolite identification software. The software facilitates automated acquisition, processing, and reporting of liquid chromatography/mass spectrometry (LC/MS) data in support of biotransformation studies. MetWorks software is compatible with all Thermo Scientific mass spectrometers capable of performing MS/MS and MS^n fragmentation. It also allows high-resolution mass data from the Thermo Scientific LTQ Orbitrap and LTQ FT instruments to be fully leveraged. The new program's enhanced tools facilitate distinguishing xenobiotic components from endogenous biological matrix interferences in LC/MS^n chromatograms and spectra. Improvements to the software include automatic generation of a data-dependent parent mass table and component filtering to exclude duplicates originating from adducts and isotopic contributions. Another improvement is the flexibility to apply up to six multiple mass defect filters, which are based on high-resolution, exact mass, and mass deficiencies of the parent drug and its putative metabolites. Chromatograms and spectra filtered with them contain fewer endogenous compounds and excipient materials, which reduces the need for manual intervention to eliminate false positives from the results.

Thermo Fisher
For information 408-965-6022
www.thermofisher.com

Proteomics Mass Spectrometry

With the addition of Compass 1.3 ion trap software, the high-capacity trap HCTultra PTM Discovery System is the first commercial mass spectrometer equipped with electron transfer dissociation (ETD) and proton transfer reaction (PTR) capabilities. The instrument features a unique and innovative ion optics design and chemistry setup for ETD/PTR that allows rapid, routine top-down characterization of large peptides and mid-size proteins in the high-capacity ion trap with superior sensitivity. The usefulness of ETD/PTR is demonstrated by the unambiguous characterization of posttranslational modifications. Compass 1.3 also introduces a novel collision-induced dissociation (CID) fragmentation mode that eliminates the low mass cut-off of ion traps in MS/MS. It enables multiplexed quantitative proteome analyses on Bruker's high-capacity ion-trap systems. A new AutoMSn mode in Compass 1.3 automatically eliminates the most abundant ions in order to further enhance dynamic range.

Bruker Daltonics
For information 978-667-9580
www.bdal.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

SCIENTIST LAWYERS
TEST TUBES TO BRIEFS

Increasingly job ads like the one to the right are targeting scientists in a broad range of fields, especially engineering, chemistry, and biotechnology. From newly minted Ph.D.s to established faculty, scientists are joining the ranks of patent agents, patent examiners, and technology transfer specialists—all careers that do not require law degrees. Those willing to go back to school to study law will have access to a wider range of opportunities, not to mention higher salaries. What these individuals bring to their jobs are strong analytical skills and an understanding of complex technologies and their applications. **By Laura Bonetta**

In the United States, scientists can join law firms fresh out of graduate school. Then, while helping with patent applications and other tasks, they typically prepare to pass an exam by the US Patent and Trademark Office (USPTO), which allows them to become patent agents.

James Dilmore went this route six years ago, when he joined the Pittsburgh-based international law firm Reed Smith. "I had found scientific research exciting but also very frustrating," says Dilmore, who obtained his Ph.D. in neuroscience. "I also wanted a position with a bit more stability than a series of postdocs."

By chance, the wife of one of his thesis committee members was getting ready to leave her post at a local law firm. She introduced Dilmore to her boss. "I did not have aspirations to go into law, but the job appealed to me," says Dilmore. He started as a technical adviser at Reed Smith in February 2001 and passed the patent exam in April 2002.

As a patent agent, Dilmore is responsible for filing patent applications on behalf of several clients. He also spends a good part of his workday advising attorneys at Reed Smith involved in litigation cases. "One of the roles I see myself filling is to ensure that lawyers do not press scientific arguments so far that the science is no longer accurate," says Dilmore. "I can recognize scientific weaknesses on either side of the argument."

And as litigations often require the testimony of renowned scientists as expert witnesses Dilmore is often asked to act as "translator." "It is part of my job to explain to scientists what is important legally," he explains. "On the other side, I help the attorneys understand what is going on with the science."

Second Degree

While it is possible to remain a patent agent, or even a technology specialist, within a law firm, a law degree is needed for career advancement. In the United States doing things like filing appeals from the USPTO to courts, negotiating licenses to use patented technology, or suing those who breach contracts requires a law degree.

Patricia Granahan started working toward a Juris Doctor (J.D.) degree right after completing her Sc.D. at the Harvard School of Public Health. "From the time I was in high school, I knew I wanted to combine an advanced degree in science with a law or business degree," she recalls. "I didn't necessarily know what I'd do with the combination."

The timing turned out to be perfect. In 1980, while Granahan was in law school, Ananda Chakrabarty was awarded a US patent for genetically engineered *Pseudomonas* he created while working for General Electric Company. The landmark case, which allowed for the first time the patenting of a living organism, paved the way for a flurry of biotechnology patents.

Granahan, who was initially hired as a part-time scientific consultant by a firm in the Boston area, soon had "more work in biotechnology patents than I could do," she recalls. Indeed the field of biotechnology patent law quickly grew, along with the demand for lawyers with advanced degrees in biology and medicine. continued »

> "You have to be a very organized person and to be able to constantly reprioritize."


Philip Webber

UPCOMING FEATURES

Careers in Preclinical Drug Discovery and Development (online only)—March 7

Focus on Asia—March 21

Vaccine Research—March 28

Postdoctoral Opportunities

"This is a career that gives you a chance to really thrive. It's not a practice where you'll ever be able to say, 'Now I know everything.'"
—Patricia Granahan

Granahan stayed with the same firm for 16 years, moving up the ranks while raising her two children. After working as an in-house lawyer at the newly established Whitehead Institute for Biomedical Research and as a partner in a Boston firm, in 2006 Granahan joined Wolf, Greenfield & Sacks, PC, where she is currently a shareholder.

Twenty-five years after starting her career in patent law, she still really enjoys what she does. "The biotech aspect of patent law has changed tremendously since I started in practice and the rules of practice have changed just as significantly. We have to keep up with changes in both the science and the law," she says. "This is a career that gives you a chance to really thrive. It's not a practice where you'll ever be able to say, 'Now I know everything.'"

Balancing School and Work

Unlike Granahan, **Karen Brown** obtained her law degree on the job. After completing a postdoc at the National Institutes of Health in the field of cell biology, Brown decided to test the waters as a patent examiner with the USPTO in Alexandria, Virginia. "I thought that if I liked it at the USPTO, it would give me an entry into patent law. If I hated it, I could go back to the bench," she recalls.

But she liked it, and three years later she started applying to law firms. One of the things she was looking for was a firm with a patent agent program in place, where the firm would pay for employees to attend law school.

In April 1998, she moved to Fish & Neave in New York City and started law school the same year, graduating four years later. "I worked all day and went to school at night," she explains. "It was brutal. I was feeling that I was behind in every aspect of my life."

After working as an associate at Fish & Neave and then in-house at a large pharmaceutical company, three years ago Brown got a position as an in-house attorney with the biotech company Vertex Pharmaceuticals in Cambridge, Massachusetts. "At the firm I was dealing with a wide variety of technologies. Now my focus is narrower but I am involved more deeply," she explains. "With every patent application, I have to consider how it fits into the intellectual property strategy of the company."

Another difference is that she does not have to worry about billing a certain number of hours each week, the typical way law firms keep track of their lawyers' accomplishments. "With two young children I did not want to be billing 60 hours a week," says Brown. "There are times when I take work home or work on weekends, but I try hard not to do that."

Wide Range of Opportunities

Erich Veitenheimer also used the USPTO as a stepping-stone to a career in patent law. A senior corn breeder with DeKalb Genetics International, he was alerted by a colleague to an ad for a job with the USPTO for a patent examiner with knowledge of plant biotechnology. "It sounded fascinating," says Veitenheimer, who was offered the job during his initial phone call inquiring about the post.

After being at the USPTO for two years Veitenheimer started attending Georgetown University Law Center in the evenings, with the USPTO helping to pay the tuition fees. "It was a tremendously long day but the stimulation was worth it," he says. "Walking into my classes I felt my brain cells tingling."

When two years later the USPTO changed its policy and reduced the amount it would pay for its examiners to attend law school, Veitenheimer became a patent agent with a mid-size patent firm in Virginia willing to support his studies. He subsequently moved to a smaller firm and then a large international firm, where he progressed all the way to partner. In 2005, he became a partner in the Washington, D.C., office of Cooley Godward Kronish, LLP, a national technology firm that serves primarily small to mid-size biotech companies.

Veitenheimer says patent law allows scientists to stay close to the science on topics ranging from stem cells, to cancer and HIV drugs, to RNA interference. "The range of topics is tremendous," says Veitenheimer. "You get to know the work intimately."

And the opportunities go beyond patent law. "About half of our lawyers do pro bono work, but not in intellectual property law," says Veitenheimer. "They come back to the firm after being in front of a judge in, for example, an immigration case and they are very excited and energized."

Boston University
www.bu.edu

Cooley Godward Kronish, LLP
www.cooley.com

DeKalb Genetics International
www.asgrowanddekalb.com

Fish & Neave (now Ropes & Gray LLP)
www.ropesgray.com

Frank B. Dehn & Co.
www.frankbdehn.com

General Electric Company
www.ge.com

Georgetown University Law Center
www.law.georgetown.edu

Harvard School of Public Health
www.hsph.harvard.edu

National Cancer Institute/NIH
www.nci.nih.gov

Reed Smith, LLP
www.reedsmith.com

USPTO
www.uspto.gov

Vertex Pharmaceuticals
www.vpharm.com

Warwick University, UK
www.warwick.ac.uk

Whitehead Institute for Biomedical Research
www.wi.mit.edu

Wolf, Greenfield & Sacks
www.wgslaw.com

Additional online resources

American Intellectual Property Law Association
www.aipla.org

European Patent Office
www.epo.org/about-us.html

UK Intellectual Property Office
www.ipo.gov.uk/patent

US Patent and Trademark Office
www.uspto.gov

Law-specific job sites

www.insidecareers.co.uk

www.ipjobs.co.uk

www.intelproplaw.com/Jobs.shtml

Patent Law in Europe

Scientists are also in high demand in law firms outside the United States. While working toward a Ph.D. at Warwick University in the UK, **Philip Webber** attended a course meant to teach chemical patent attorneys about biotechnology. "I did not know anything about patent law at that time. It was purely fortuitous that my university was running these kinds of courses and that I attended one of them," he recalls.

The course sparked an interest in the subject. Webber, who had already decided he did not want to continue with bench research, started sending out applications to several law firms. He was hired by the London-based Frank B. Dehn & Co. in the spring of 1992.

A year after joining his firm, Webber passed a set of exams, called "foundation papers," which test understanding of basic concepts such as patents, trademarks, and copyright. He then passed two additional sets of longer and more challenging exams that qualified him to practice patent law within the British and European patent offices.

However, because UK patent attorneys, unlike their American counterparts, do not have law degrees, they have limited rights to go in front of a judge to handle, for example, patent litigation. Such cases are generally handled by solicitors and barristers, who are less likely to have formal scientific training. British intellectual property solicitors typically practice in all areas of technology with biotech being only a small portion of it.

Regardless of the country of practice, a career in patent law requires good analytical and communication skills. "The ability to write clearly and concisely is paramount," says Webber. "Often we give job applicants a test where they are given something simple like a desk staple remover and told to describe it in everyday language."

Another important skill is to keep all the balls in the air at the same time. "At any one time I have 50 or so patent applications going through the exam process. A lot of the deadlines are set in stone. And I am bombarded with e-mails from clients all the time," says Webber. "You have to be a very organized person and to be able to constantly reprioritize."

Good people skills are also helpful as patent agents and attorneys deal with a variety of people on a daily basis, from scientists to business executives, and patent examiners to other lawyers.

Related Careers

Patent law is not the only career available to scientists interested in intellectual property issues. Some scientists land jobs at the patent office and review patent applications for a living. Others find satisfying careers as technology transfer officers at universities or research institutes. **Sean Lee** joined the tech transfer office at Boston University after having worked at two startup companies. "There is so much going on in technology transfer," says Lee. "For me it was like a kid walking into a candy store."

Lee is the first point of contact for physical scientists who have made a discovery that may have commercial value. "In technology transfer you need to look at the world a bit differently than when you are doing research," says Lee. "You don't want to ask what is most interesting, but rather who really cares about this and is going to pay for it."

Lee's experience is not unlike that of **Lisa Finkelstein,** who joined the technology transfer office at the National Cancer Institute (NCI) of the US National Institutes of Health (NIH) in September 2005. She acts as a conduit for scientists at several NCI labs to assist with submitting internal paperwork on inventions. "Anyone at NCI can contact our office for guidance and advice on potentially patentable new discoveries," she says. She also drafts material transfer and collaboration agreements for the labs on her docket.

Some of her colleagues moved on to technology transfer offices in universities or were hired by pharmaceutical companies or law firms. "Once you get the training here, you are pretty much set for wherever you want to go," she says. "There are a lot of options."

Regardless of the path they take, scientists who have pursued intellectual property law as a career say it nourishes their love for science. "It is a great fit for people who enjoy science and want to have the ability to read about it, talk about it, and be involved in it, but without actually doing it," says Finkelstein.

"In technology transfer you need to look at the world a bit differently than when you are doing research."
—Sean Lee

Finding a Job in Patent Law

Although scientists are in high demand in law firms, the competition for a job is stiff. An ad typically gets 100 to 200 applications. For applicants to stand out, it is important that they have not only good scientific credentials, but also a demonstrated interest in intellectual property.

Scientists who have become patent agents and lawyers advise postdocs and students interested in a career in patent law to take a look at some patent applications; these are freely downloadable from patent office websites. In addition, they should take advantage of courses in intellectual property law being offered at many universities. Some technology transfer offices at universities also offer internships or volunteer work. Finally, many major cities have an intellectual property association that provides opportunities for networking.

Laura Bonetta is a freelance writer based in the Washington, D.C., area.

DOI: 10.1126/science.opms.r0800048